



**IntechOpen**

# Applications of Monte Carlo Method in Science and Engineering

*Edited by Shaul Mordechai*





---

# **APPLICATIONS OF MONTE CARLO METHOD IN SCIENCE AND ENGINEERING**

---

Edited by **Shlomo Mark** and **Shaul Mordechai**

## Applications of Monte Carlo Method in Science and Engineering

<http://dx.doi.org/10.5772/1954>

Edited by Shaul Mordechai

### Contributors

Lester Alfonso, Graciela Raga, Darrel Baumgardner, Kunihiro Suzuki, Yu-Xuan Ren, Jian-Guang Wu, Yin-Mei Li, Dengming Xiao, Vera Behar, Christo Kabakchiev, Ivan Garvanov, Hermann Rohling, Rita Khanna, Sven K. Esche, Tran Nguyen Lan, Tran Hoang Hai, Benjamin Peter Geisler, Laszlo Baranyai, Ming Wu, Craig A. Kruschwitz, Jianye Ching, Hossein Haroonabadi, Fengmin Wu, Juanmei Hu, Antonio Martinez, Dolores Rodrigo, Fernando Sampedro, M. Consuelo Pina, Javier Collado, Antonio Falcó, Fan Zhong, Frank Sukowski, Norman Uhlmann, Soumen Kumar Roy, Nababrata Ghoshal, Kisor Mukhopadhyay, Shyamal Bhar, Atsushi Mori, Qing Hao, Gang Chen, Ali Asghar Mowlavi, Mario de Denaro, Maria Rosa Fornasier, Margherita Premuda, Thejappa Golla, Sayyad Zahid Qamar, Anwar Khalil Sheikh, Abul Fazal M. Arif, Tasneem Pervez, Dôme Tanguy, Tomas Gonzalez, Javier Mateos, Ignacio Iñiguez-de-la-Torre, Beatriz Garcia Vasallo, Helena Rodilla, Miroslav Morhac, Eva Morhacova, Dominique Persano Adorno, Sorina Camarasu-Pop, Tristan Glatard, Hugues Benoit-Cattin, David Sarrut, Bart Verberck, Mamdouh A. Al-Harhi, Kalliopi Trohidou, Marianna Vasilakaki, Katsuhiko Yamaguchi, Kenji Suzuki, Osamu Nittono, Marie-Anne Jaud, Sylvain Barraud, Jérôme Saint Martin, Arnaud Bournel, Philippe Dollfus, Hervé Jaouen, Guillermo Albareda, Fabio Traversa, Abdelilah Benali, Xavier Oriols, Theodore A. Steriotis, Maria Konstantakou, Anastasios Gotzias, Michael Kainourgiakis, Athanasios Stubos, A Slabi, Vladimir Stary, Zejun Ding, Yonggang Li, Shifeng Mao, Dragica Vasileska

### © The Editor(s) and the Author(s) 2011

The moral rights of the and the author(s) have been asserted.

All rights to the book as a whole are reserved by INTECH. The book as a whole (compilation) cannot be reproduced, distributed or used for commercial or non-commercial purposes without INTECH's written permission.

Enquiries concerning the use of the book should be directed to INTECH rights and permissions department ([permissions@intechopen.com](mailto:permissions@intechopen.com)).

Violations are liable to prosecution under the governing Copyright Law.



Individual chapters of this publication are distributed under the terms of the Creative Commons Attribution 3.0 Unported License which permits commercial use, distribution and reproduction of the individual chapters, provided the original author(s) and source publication are appropriately acknowledged. If so indicated, certain images may not be included under the Creative Commons license. In such cases users will need to obtain permission from the license holder to reproduce the material. More details and guidelines concerning content reuse and adaptation can be found at <http://www.intechopen.com/copyright-policy.html>.

### Notice

Statements and opinions expressed in the chapters are those of the individual contributors and not necessarily those of the editors or publisher. No responsibility is accepted for the accuracy of information contained in the published chapters. The publisher assumes no responsibility for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained in the book.

First published in Croatia, 2011 by INTECH d.o.o.

eBook (PDF) Published by IN TECH d.o.o.

Place and year of publication of eBook (PDF): Rijeka, 2019.

IntechOpen is the global imprint of IN TECH d.o.o.

Printed in Croatia

Legal deposit, Croatia: National and University Library in Zagreb

Additional hard and PDF copies can be obtained from [orders@intechopen.com](mailto:orders@intechopen.com)

Applications of Monte Carlo Method in Science and Engineering

Edited by Shaul Mordechai

p. cm.

ISBN 978-953-307-691-1

eBook (PDF) ISBN 978-953-51-5604-8

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,100+

Open access books available

116,000+

International authors and editors

120M+

Downloads

151

Countries delivered to

Our authors are among the  
Top 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)





# Meet the editors



Dr. Shlomo Mark (PhD) is the Chair of the Software Engineering Department and the head of the Negev Monte Carlo Research Center (NMCRC) at SCE – Sami Shamoon College of Engineering, Israel. Dr. Mark earned his PhD in 2003 from the department of nuclear engineering, Ben Gurion University of the Negev, Israel, in the field of Experimental Methods and Computational Modeling for Physical and Medical Applications. His main research interests are: Monte Carlo Simulations, Scientific Programming in Software Engineering, Developing methods and tools for Testing and Verification of Scientific and Monte Carlo Codes, Service Science, Agile Programming, Computational Modeling for Physical, Environmental and Medical Applications. He has published many papers in peer review journals in the field of Monte Carlo Simulation and Computational Modeling.



Dr. Shaul Mordechai (PhD) is a Professor of Physics and Head of the Biomedical Spectroscopy Laboratory at the Department of Physics, Ben Gurion University, Israel. He received his PhD from the Racah Institute of Physics, Hebrew University Jerusalem. His research interest include: Medical Physics, Biomedical Optics, FTIR-Microscopy, FTIR-Imaging, Mid-IR Spectrometry, Fluorescence Spectroscopy, Optical Diagnostics, and Biomedical Applications of Monte Carlo Simulations. He was a Visiting Scientist at the University of Texas at Austin, Los Alamos National Laboratory and the University of Pennsylvania. He is a Member of the Editorial Board of World Journal of Biological Chemistry (WJBC), Associate Editor - Medical Physics, Fellow of the American Society for Laser Medicine and Surgery. He has coauthored many papers in Biomedical Optics, Tissue Microscopy, Optical Diagnostics and Applications of Monte Carlo Simulations in Biomedical Optics.



---

# Contents

---

## **Preface XIII**

- Chapter 1 **Monte Carlo Simulations in NDT 1**  
Frank Sukowski and Norman Uhlmann
- Chapter 2 **Application of Monte Carlo Simulation  
in Optical Tweezers 21**  
Yu-Xuan Ren, Jian-Guang Wu and Yin-Mei Li
- Chapter 3 **Enabling Grids for GATE Monte-Carlo Radiation  
Therapy Simulations with the GATE-Lab 35**  
Sorina Camarasu-Pop, Tristan Glatard,  
Hugues Benoit-Cattin and David Sarrut
- Chapter 4 **Monte Carlo Simulation for Ion Implantation Profiles,  
Amorphous Layer Thickness Formed by the Ion  
Implantation, and Database Based on Pearson Function 51**  
Kunihiro Suzuki
- Chapter 5 **Application of Monte Carlo Simulation  
in Industrial Microbiological Exposure Assessment 83**  
Javier Collado, Antonio Falcó, Dolores Rodrigo,  
Fernando Sampedro, M. Consuelo Pina and Antonio Martínez
- Chapter 6 **Monte Carlo Simulation of Radiative Transfer  
in Atmospheric Environments for Problems Arising  
from Remote Sensing Measurements 95**  
Margherita Premuda
- Chapter 7 **Monte Carlo Simulation of Pile-up Effect  
in Gamma Spectroscopy 125**  
Ali Asghar Mowlavi, Mario de Denaro and Maria Rosa Fornasier
- Chapter 8 **Monte Carlo Simulations of Microchannel  
Plate-Based, Time-Gated X-ray Imagers 141**  
Craig A. Kruschwitz and Ming Wu

- Chapter 9 **Many-particle Monte Carlo Approach to Electron Transport** 167  
G. Albareda, F. L. Traversa, A. Benali and X. Oriols
- Chapter 10 **Monte-Carlo Simulation in Electron Microscopy and Spectroscopy** 195  
Vladimír Starý
- Chapter 11 **Monte Carlo Simulation of SEM and SAM Images** 231  
Y.G. Li, S.F. Mao and Z.J. Ding
- Chapter 12 **Monte Carlo Simulation of Insulating Gas Avalanche Development** 297  
Dengming Xiao
- Chapter 13 **Monte Carlo Simulation of Electron Dynamics in Doped Semiconductors Driven by Electric Fields: Harmonic Generation, Hot-Carrier Noise and Spin Relaxation** 331  
Dominique Persano Adorno
- Chapter 14 **A Pearson Effective Potential for Monte-Carlo Simulation of Quantum Confinement Effects in nMOSFETs** 359  
Marie-Anne Jaud, Sylvain Barraud, Philippe Dollfus, Jérôme Saint-Martin, Arnaud Bournel and Hervé Jaouen
- Chapter 15 **Monte Carlo Device Simulations** 385  
Dragica Vasileska, Katerina Raleva and Stephen M. Goodnick
- Chapter 16 **Wang-Landau Algorithm and its Implementation for the Determination of Joint Density of States in Continuous Spin Models** 431  
Soumen Kumar Roy, Kisor Mukhopadhyay, Nababrata Ghoshal and Shyamal Bhar
- Chapter 17 **Characterizing Molecular Rotations using Monte Carlo Simulations** 451  
Bart Verberck
- Chapter 18 **Finite-time Scaling and its Applications to Continuous Phase Transitions** 469  
Fan Zhong
- Chapter 19 **Using Monte Carlo Method to Study Magnetic Properties of Frozen Ferrofluid** 495  
Tran Nguyen Lan and Tran Hoang Hai
- Chapter 20 **Monte Carlo Studies of Magnetic Nanoparticles** 513  
K. Trohidou and M. Vasilakaki

- Chapter 21 **Monte Carlo Simulation for Magnetic Domain Structure and Hysteresis Properties** 539  
Katsuhiko Yamaguchi, Kenji Suzuki and Osamu Nittono
- Chapter 22 **Monte Carlo Simulations of Grain Growth in Polycrystalline Materials Using Potts Model** 563  
Miroslav Morháč and Eva Morháčová
- Chapter 23 **Monte Carlo Simulations of Grain Growth in Metals** 581  
Sven K. Esche
- Chapter 24 **Monte Carlo Simulations on Defects in Hard-Sphere Crystals Under Gravity** 611  
Atsushi Mori
- Chapter 25 **Atomistic Monte Carlo Simulations in Steelmaking: High Temperature Carburization and Decarburization of Molten Steel** 629  
R. Khanna, R. Mahjoub and V. Sahajwalla
- Chapter 26 **GCMC Simulations of Gas Adsorption in Carbon Pore Structures** 653  
Maria Konstantakou, Anastasios Gotzias, Michael Kainourgiakis, Athanasios K. Stubos and Theodore A. Steriotis
- Chapter 27 **Effect of the Repulsive Interactions on the Nucleation and Island Growth: Kinetic Monte Carlo Simulations** 677  
Hu Juanmei and Wu Fengmin
- Chapter 28 **Monte Carlo Methodology for Grand Canonical Simulations of Vacancies at Crystalline Defects** 687  
Dôme Tanguy
- Chapter 29 **Frequency-Dependent Monte Carlo Simulations of Phonon Transport in Nanostructures** 707  
Qing Hao and Gang Chen
- Chapter 30 **Performance Analysis of Adaptive GPS Signal Detection in Urban Interference Environment using the Monte Carlo Approach** 735  
V. Behar, Ch. Kabakchiev, I. Garvanov and H. Rohling
- Chapter 31 **Practical Monte Carlo Based Reliability Analysis and Design Methods for Geotechnical Problems** 757  
Jianye Ching
- Chapter 32 **A Monte Carlo Framework to Simulate Multicomponent Droplet Growth by Stochastic Coalescence** 781  
Lester Alfonso, Graciela Raga and Darrel Baumgardner

- Chapter 33 **Monte Carlo Simulation of Room Temperature Ballistic Nanodevices 803**  
Ignacio Íñiguez-de-la-Torre, Tomás González,  
Helena Rodilla, Beatriz G. Vasallo and Javier Mateos
- Chapter 34 **Estimation of Optical Properties in Postharvest and Processing Technology 829**  
László Baranyai
- Chapter 35 **MATLAB Programming of Polymerization Processes using Monte Carlo Techniques 841**  
Mamdouh A. Al-Harhi
- Chapter 36 **Monte Carlo Simulations in Solar Radio Astronomy 857**  
G. Thejappa and R. J. MacDowall
- Chapter 37 **Using Monte Carlo Simulation for Prediction of Tool Life 881**  
Sayyad Zahid Qamar, Anwar Khalil Sheikh,  
Tasneem Pervez and Abul Fazal M. Arif
- Chapter 38 **Loss of Load Expectation Assessment in Electricity Markets using Monte Carlo Simulation and Neuro-Fuzzy Systems 901**  
H. Haroonabadi
- Chapter 39 **Automating First- and Second-order Monte Carlo Simulations for Markov Models in TreeAge Pro 917**  
Benjamin P. Geisler
- Chapter 40 **Monte Carlo Simulations of Adsorbed Molecules on Ionic Surfaces 931**  
Abdulwahab Khalil Sallabi

---

## Preface

---

**Monte Carlo simulation**, the iterative computational method used to examine and investigate the behavior of physical and mathematical systems utilizing stochastic techniques. It is a widely used method and a successful statistical tool in studying a broad array of problems, areas and cases in which it is infeasible or impossible to compute exact results utilizing deterministic algorithms.

The Monte Carlo method has proven to be a very useful statistical sampling computational technique in attaining approximate numerical solutions to system and quantitative problems which are complex, nonlinear, involve uncertain parameters, and that are otherwise too complicated to solve analytically. In such areas of problem solving, when compared to other methods of analysis, Monte Carlo approaches are known to be the most accurate. Historically, however, because of the relatively large amount of computational time required, these techniques were considered fairly burdensome. Nowadays, as a result of the ever-increasing computing power, as well as the increasing availability of distributed resources, these computations can be substantially accelerated.

In today's world, with the wide prevalence of novel programming languages and tools, the rapid growth of computing power and the availability of ever more advanced and powerful hardware, the need for increasingly complex and powerful computational solutions such as Monte Carlo simulation and applications is growing exponentially. The utilization of Monte Carlo methods, simulations and applications, is found in widely disparate fields and areas of application such as nuclear physics, reliability, networks, finance and business, engineering, economics, risk analysis, project management, the study of heat transfer, molecular dynamic analysis, environmental sciences, chemistry, telecommunications, engineering, games and so forth.

In this book, *Applications of Monte Carlo Method in Science and Engineering*, we further expose the broad range of applications of Monte Carlo simulation in the fields of Quantum Physics, Statistical Physics, Reliability, Medical Physics, Polycrystalline Materials, Ising Model, Chemistry, Agriculture, Food Processing, X-ray Imaging, Electron Dynamics in Doped Semiconductors, Metallurgy, Remote Sensing

and much more diverse topics. The book chapters included in this volume clearly reflect the current scientific importance of Monte Carlo techniques in various fields of research.

**Shlomo Mark**

Negev Monte Carlo Research Center and Department of Software Engineering,  
SCE - Sami Shamoon College of Engineering, Bialik/Basel Sts. Beer Sheva 84100,  
Israel

**Shaul Mordechai**

Department of Physics and the Cancer Research Center,  
Ben-Gurion University (BGU), Beer-Sheva, 84105,  
Israel

# Monte Carlo Simulations in NDT

Frank Sukowski and Norman Uhlmann

*Fraunhofer Institute for Integrated Circuits IIS, Development Center X-ray Technology  
(EZRT)  
Germany*

## 1. Introduction

X-ray techniques are commonly used in the fields of non-destructive testing (NDT) of industrial parts, material characterization, security and examination of various other specimens. The most used techniques for obtaining images are radioscopy for 2D and computed tomography (CT) for 3D imaging. Apart from these two imaging techniques, where X-ray radiation penetrates matter, other methods like refraction or fluorescence analysis can also be used to obtain information about objects and materials. The vast diversity of possible specimen and examination tasks makes the development of universal X-ray devices impossible. It rather is necessary to develop and optimize X-ray machines for a specific task or at least for a limited range of tasks. The most important parameters that can be derived from object geometry and material composition are the X-ray energy or spectrum, the dimensions, the examination geometries and the size of the detector. The task itself demands a certain image quality which depends also on the X-ray spectrum, the examination geometry and furthermore on the size of the X-ray source's focal spot and the resolution of the detector.

Monte-Carlo (MC) simulations are a powerful tool to optimize an X-ray machine and its key components. The most important components are the radiation source, e.g. an X-ray tube and the detector. MC particle physics simulation codes like EGS (Nelson et al., 1985) or GEANT (Agostinelli et al., 2003) can describe all interactions of particles with matter in an X-ray environment very well. Almost all effects can be derived from these particle physics processes. The MC codes are event based. Every single primary particle is generated and tracked along with all secondary particles until the energy of all particles drops below a certain threshold. The primaries are generated one after another, since no interactions between particles take place.

When simulating X-ray sources, in most cases X-ray tubes, the primary particles are electrons. The electron beam is parameterized by the electrons' kinetic energy and the intensity profile along the cross-section of the beam. When hitting the target, X-rays are generated by interaction of electrons with the medium. The relevant magnitudes for imaging are the X-ray energy spectrum and the effective optical focal spot size (Morneburg, 1995).

The most used imaging systems in the field of NDT are flat panel detectors. There are two basic types of detectors: Direct converting semiconductor detectors and indirect converting scintillation detectors. The type of particle interactions in the respective sensor layer determines the detection efficiency and effective spatial resolution. Interaction of X-rays in direct converting detectors produces electron-hole-pairs in the semiconductor materials. The free charge carriers drift to electrodes, where the current can be measured. MC simulations can

describe the X-ray absorption and scattering as well as the electron drift which leads to image blurring. Measuring X-rays with scintillation detectors works differently. X-rays interact in the scintillation layer and produce visible photons, which are detected in a CCD or CMOS chip. In addition to X-ray scattering and electron drift the diffusion of the visible photons in the scintillation layer contributes greatly to image blurring (Beutel et al., 2000). In any case, a thicker sensor layer improves the detection efficiency on the one hand, which leads to shorter measurement times, but decreases the spatial resolution on the other hand. Finding the optimal trade-off between efficiency and resolution by designing detector properties is an excellent task for MC simulations.

Another application field of MC simulations are feasibility studies for special examination tasks in order to evaluate physical limits of different imaging methods. These studies are not limited to radioscopic methods, but include other ways to obtain information about specimens like refractive, diffractive and backscatter imaging as well as fluorescence analysis and many more.

In this chapter MC applications aimed at the optimization of X-ray setups for specific tasks and feasibility studies are introduced.

The used Monte-Carlo code is called ROSI (ROentgen Simulation), which was developed by J. Giersch and A. Weidemann at the University of Erlangen (Giersch et al., 2003). It is an object oriented program code and the simulation runs can be parallelized in a computer network for largely increasing the performance. It is based on the particle physics codes EGS4 for general electromagnetic particle interactions and LSCAT for low energy processes.

## 2. Simulation of X-ray sources

### 2.1 X-ray source characteristics in NDT imaging

In common X-ray tubes, radiation is produced by accelerating electrons via a potential difference between the cathode (the electron emitter) and the anode (the X-ray target). When the electrons hit the target, they are decelerated hard by collisions with electrons of the target material or in the coulomb fields of atomic cores. X-ray radiation is produced in two different processes. Since electrons are charged, acceleration or in this case deceleration can cause emission of photons. The energy of these photons corresponds to the electrons' energy loss during the deceleration process, so the maximum possible energy corresponds to the acceleration voltage ( $E_{\max} = e \cdot U$ ). This process is called bremsstrahlung. The other process is called characteristic or fluorescence radiation and takes place when electrons ionize the target material by hitting bound electrons. The excited atoms change into their ground state very quickly by electronic transition from a high to the lower vacant energy level. During this process a photon is emitted, whose energy corresponds to the difference in these energy levels (Morneburg, 1995).

#### 2.1.1 Energy spectrum

In the field of X-ray imaging the kind of application forces all necessary source properties. When penetration techniques like radiography or computed tomography are used, the X-ray radiation energy is one of the most important parameters. The radiation must partially penetrate the object to obtain the highest possible contrast between high and low absorbing parts of the specimen. With X-ray tubes as sources, the energy spectrum can be shaped by adjusting the tube voltage and using various prefilters. Figure 1 shows spectra between 30 and 450 kV with several prefilters.

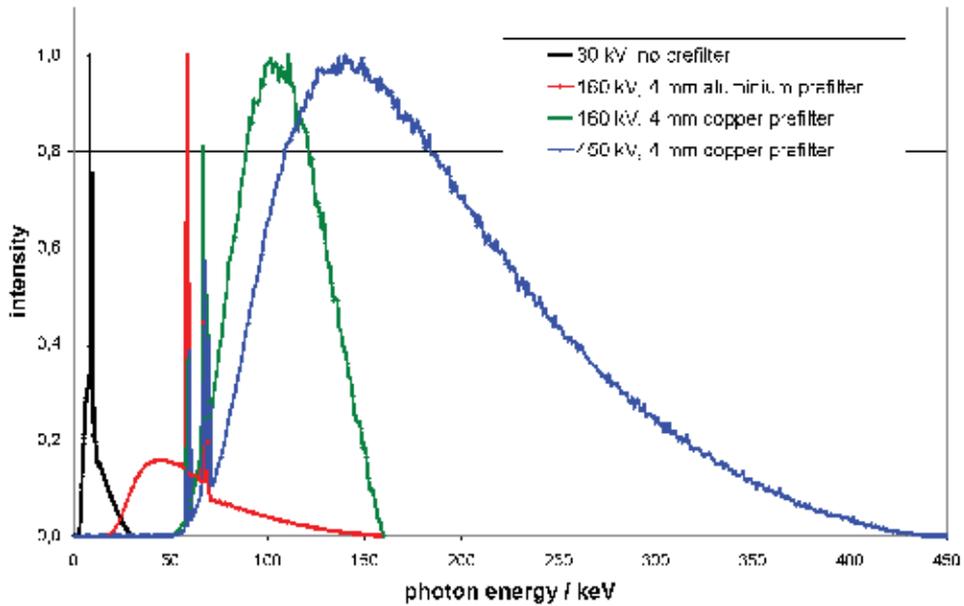


Fig. 1. X-ray spectra, normalized to a maximum of 1

The influence of the image quality can clearly be seen in 2. A Siemensstern with 8 mm thick iron and copper sections is radiographed. (a) The energy of the X-rays is not sufficient to penetrate any material, the area behind the object is completely dark. (b) The area behind the object is still very dark compared to the uncovered area, although a faint contrast between copper (darker) and iron (lighter) can be seen. Many low energy photons enhance the brightness in the uncovered area, while they are completely absorbed in the object. (c) The low-energy photons are filtered out by the prefilter and don't contribute to either the uncovered or covered image parts. The difference between these areas is reduced, while the contrast is enhanced. This spectrum would be a good choice for separating the iron and copper sections. (d) The vast majority of the photons penetrate the object regardless of the material. The complete object appears brighter, but the contrast between iron and copper is reduced again.

### 2.1.2 Focal spot size

The focal spot size  $U_F$  of the X-ray source is also a very important magnitude and has a large influence on the spatial resolution of the image, especially when working with high magnifications  $M$ . The magnification is given by the fraction of the focus-detector-distance  $FDD$  and the focus-object-distance  $FOD$ . As illustrated in 3, the geometrical unsharpness  $U_g$  of the image is given by

$$U_g = U_F (1 - M) = d \left( 1 - \frac{FDD}{FOD} \right). \quad (1)$$

### 2.1.3 Intensity

With many applications, the measurement time is crucial and should be as short as possible. The image noise on one hand results from electronic noise in detector systems, but the main

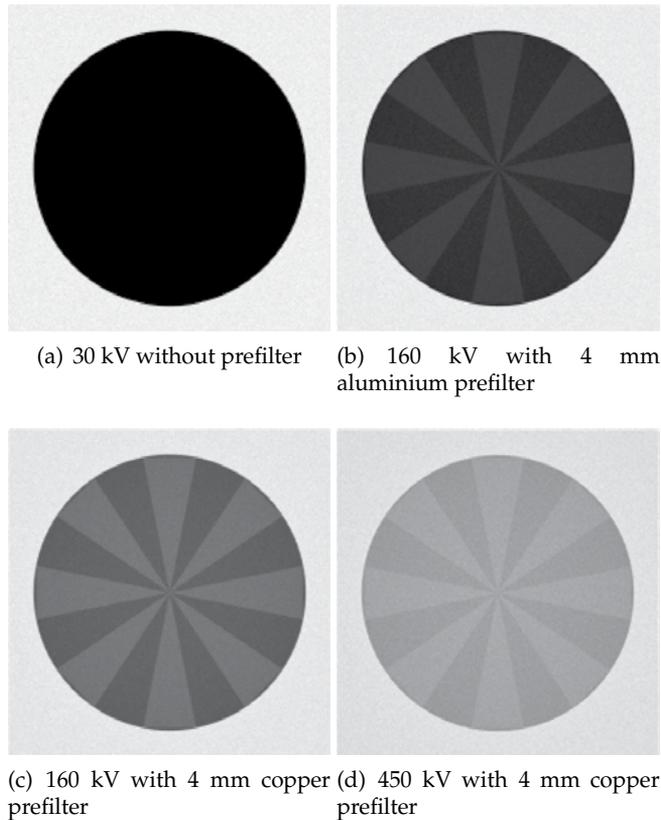


Fig. 2. Images of a Siemensstern. The sections are iron and copper with thickness of 8 mm each

part originates from poisson noise due to limited quantum statistics. Poisson noise is  $1/\sqrt{N_p}$ , where  $N_p$  is the number of events per pixel in one image. For obtaining low-noise images in a short time, the source intensity must be maximized. The number of emitted photons from an X-ray source first depends on the tube voltage  $U$ . The intensity is roughly proportional to the squared voltage. Since the voltage shapes the energy spectrum, it is not always desirable to change it for a given application. The second way to increase the intensity is to increase the tube current  $I$ , which is proportional to the intensity. The electrical power  $P$  applied to the X-ray target is  $P = U \cdot I$ . Unfortunately only about 1% of the electrical power is converted to X-rays. The vast majority of the electrical power heats up the target, which forces a limitation in the applicable current. Monte-Carlo simulations can help a great deal to optimize target material composition and geometry to increase the load capacity of targets or increase the X-ray conversion efficiency.

## 2.2 High resolution imaging

As mentioned in the above section, a small focal spot is crucial to achieve a good spatial resolution when working with high magnifications. High resolution in X-ray imaging means resolution of object details below 1 micron. For those applications, microfocus X-ray tubes with transmission targets are commonly used where the target is also the excitation window of

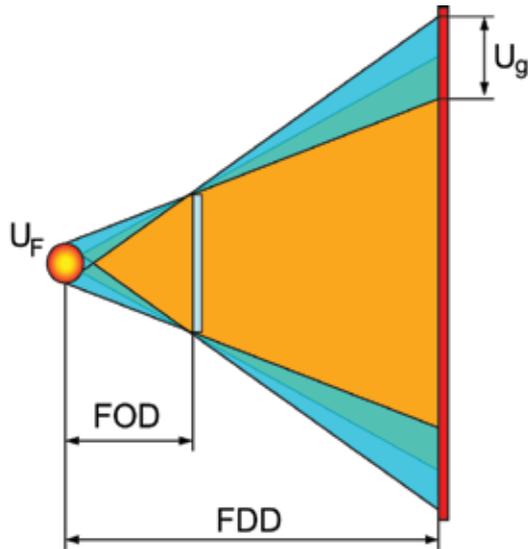


Fig. 3. Geometrical unsharpness due to X-ray source dimension

the tube. The transmission target has a great advantage since the specimen can be placed very close to the focal spot in order to achieve high magnifications. The electron beam in the X-ray tube is focused onto the target by electronic lenses. The diameter of the beam on the target surface reaches from 200 nm to several  $\mu\text{m}$  and mostly determines the X-ray focal spot size. But the diffusion of the electrons in the target, which depends largely on the target materials and layer composition can further increase the focal spot size as shown in 4. To design a target for smallest possible focal spots, Monte-Carlo simulations of electronic diffusion and X-ray production processes were performed.

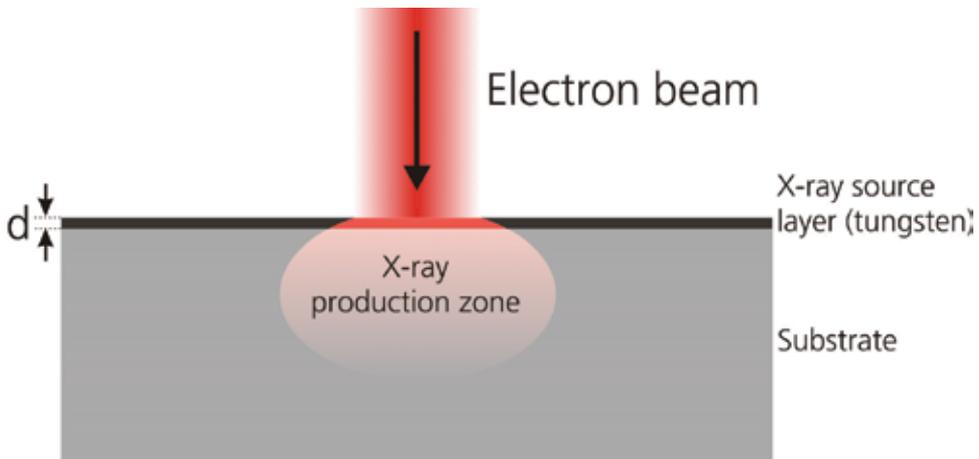


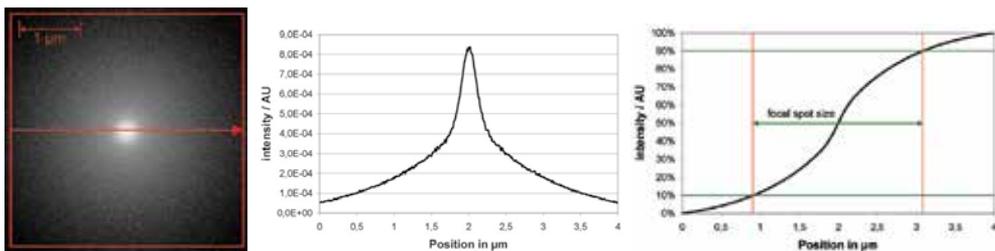
Fig. 4. Geometrical setup of a transmission target

In the simulation a parallel electron beam with electron kinetic energy between 30 and 120 keV was modeled with a gaussian intensity cross-section in both dimensions. The FWHM value of the gaussian distribution was 200 nm. The first layer material of the transmission target

is tungsten. Since the X-ray productivity rises with the atomic number proportional to  $Z^2$ , tungsten with  $Z = 74$  is a good choice. It has even more advantages, a very high melting point at over 3000 °C, a fair thermal conductivity, mechanical and chemical stability. The X-rays are produced mainly in the tungsten layer, which is also called the X-ray production layer. In the simulations, the thickness of this layer was varied from 0.05 to 7 microns (depending on electron energy). From their point of origin the photons have to pass the remaining target material to reach the side opposite the electron beam. Therefore the substrate material must fulfill several requirements. The atomic number must be quite low, so the X-rays can pass that layer without being absorbed, even at low energies. Furthermore, the substrate must have a good thermal conductivity and a high melting point so that the heat that is generated in the tungsten layer can be conducted to the air side of the target, where it can be cooled by fans for example. A performance number can be approximated by the product of thermal conductivity  $\lambda$  and maximum allowable temperature  $T_{\max}$ . A further task of the substrate is to form a mechanical closure of the vacuum vessel against the air pressure. Since the target must be thin for X-ray transmissibility, the material must be quite stable. Common materials for this task are beryllium, aluminium, diamond or other carbon configurations. The simulations were done for a 300 micron thick beryllium substrate, which forms a quite stable vacuum closure. As simulation results the diameter of the effective focal spot  $U_F$ , i.e. the area where photons are produced and the X-ray production efficiency were obtained. The total X-ray intensity  $\phi$  and the brilliance  $b$ , which is defined as the intensity divided by the source area are also important magnitudes for some applications.

$$b = \frac{\phi}{A_F} = \frac{4\phi}{\pi U_F^2} \quad (2)$$

Determining the focal spot size  $U_F$  from simulation data is shown in 5. The two-dimensional energy distribution of generated X-rays on the target was calculated with ROSI (a). The focal spot profile was taken from a line profile averaged over the whole target width in one direction (b). This profile was integrated after normalizing the total X-ray power to a value of 1. The focal spot is defined as the area where the integral value is between 0.1 and 0.9 (c).



(a) X-ray energy distribution of all generation locations (b) One-dimensional focal spot profile averaged over whole width (c) Integral over normalized profile

Fig. 5. Determination of focal spot sizes

In figure 6 the effective focal spot size  $U_F$  (a), the X-ray intensity  $\phi$  (b) and the brilliance  $b$  (c) is shown for several tungsten layer thicknesses and the tube voltages of 30, 70 and 120 kV. The intensities are calculated per target current.

For each voltage, all curves follow a similar course. The focal spot size can never be smaller than the diameter of the electron beam, so it is nearly 200 microns in diameter with very thin

tungsten layers, since only a few electrons interact with that layer and are barely scattered to distant parts of the tungsten. Due to the small interaction probability, the X-ray intensity is also very low. With thicker tungsten layers, the interaction probability and therefore the production rate of photons rises rapidly. Since the average scattering angles are quite small, especially at higher voltages, the electron beam barely broadens in that layer, keeping the focal spot size almost constant. The brilliance rises to a maximum until the tungsten becomes thick enough so that electrons can be scattered multiply, reaching distant parts of that layer, where they also produce X-rays. The result is an increase of the focal spot size. The total number of photons produced and reaching the opposite side of the target still rises until the tungsten becomes so thick, that the photons are reabsorbed by the tungsten. The focal spot size gets into saturation and the intensity is again reduced by higher target self-absorption.

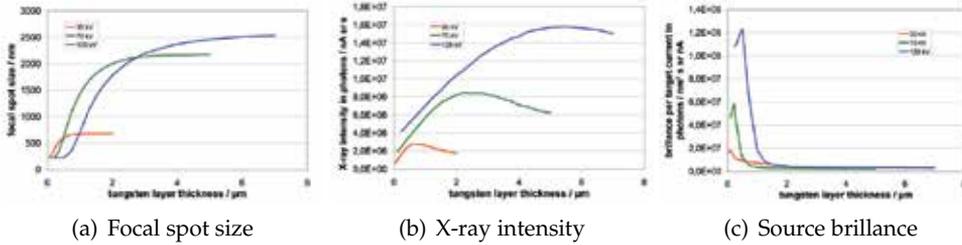


Fig. 6. Optimization of target configuration with nano focus sources

Of course the simulations can also be done with other substrate materials and thicknesses to find optimal parameters for a specific application. The Monte Carlo simulation can also calculate the heat deposition in the target volume. The data can then be taken into a heat transfer simulation tool to calculate the heat load capacity of the whole target.

### 2.3 High energy imaging

Imaging of very large and dense objects such as freight containers, whole cars (especially engines) or parts from shipbuilding requires very high energetic radiation in the MeV range to penetrate these objects. X-ray tubes on the market are available up to voltages of 450 kV, which is by far not enough. To produce high energy X-rays linear accelerators (LINACs) are commonly used. The principle in generating X-rays is the same, but the method of accelerating the electrons differs from X-ray tubes. The electrons are emitted by a gun and accelerated by bundles in a waveguide through several copper cavities. A high voltage microwave signal is applied, which accelerates the electron bundles over several cavities up to kinetic energies of some MeVs.

When electrons hit the target at these energies, X-ray radiation is almost solely produced in the direction of the impacting electrons, so X-ray targets work exclusively as transmission targets. The relativistic Lamor formula describes the angular distribution of bremsstrahlung generation (Jackson, 2006):

$$\frac{dP}{d\Omega} = \frac{e^2 \dot{v}^2}{4\pi c^3} \frac{\sin^2 \theta}{(1 - \beta \cos \theta)^5} \quad (3)$$

At very high energies and small angles,  $\beta = v/c \approx 1$ , the denominator decreases with a power of five and the whole term gets very large. Using high energy X-rays for imaging means that the radiation field is limited or at least decreases rapidly in intensity at the borders. To

choose appropriate radiation geometries for different object sizes, the radiation field has to be calculated and taken into account.

We modeled a commonly X-ray target made of 800  $\mu\text{m}$  copper and 450  $\mu\text{m}$  tungsten. The electron beam was modeled as a parallel and monoenergetic beam. The intensity cross-section was gaussian in shape with a FWHM value of 1 mm. We calculated the angular X-ray intensity distribution for energies from 1 to 18 MeV (see 7).

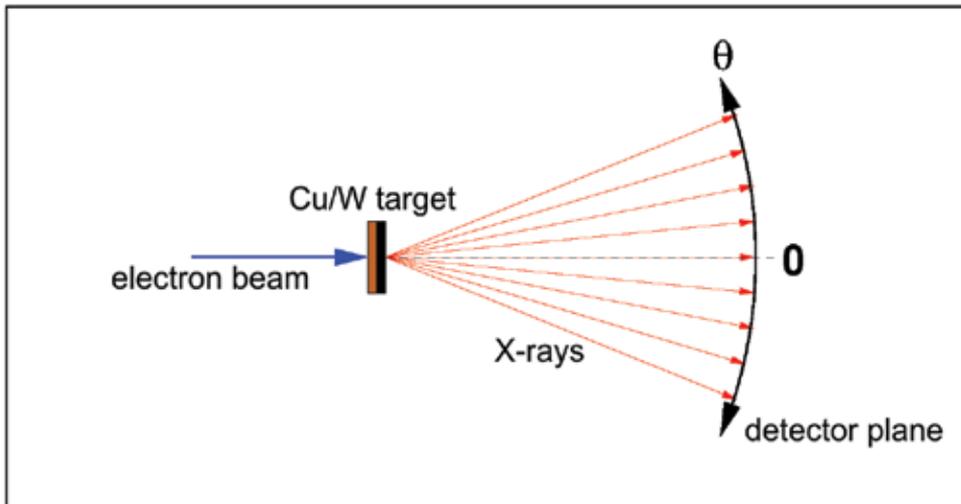
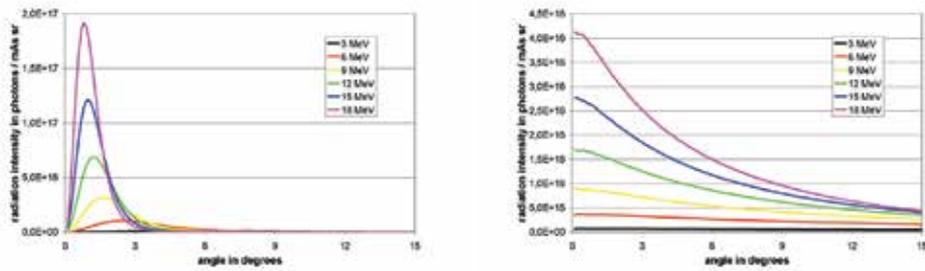


Fig. 7. Simulation setup for X-ray generation with a LINAC target

The results are shown in 8. The theoretically calculated distribution looks quite different to the simulation results. The Lamor formula assumes all electrons travelling in the forward direction ( $\theta = 0^\circ$ ) when generating bremsstrahlung. In reality the electrons can be scattered by collisions with other electrons and atomic cores while changing their direction before generating bremsstrahlung. The forward peak is blurred to higher angles. The absolute intensity increase with electron kinetic energy is described very well and corresponds to the theory.

## 2.4 Efficiency optimization

Some applications get along without high resolution or high energy sources. Sometimes a short measurement time is most essential. Inspection systems within an industrial production line have to measure prefabricated parts within a production cycle. When inspecting parts with computed tomography for reconstructing the whole 3-dimensional volume, this task is quite demanding, since the parts must be radiographed from several hundred points of view in a short time. The most important component to achieve this is a highly intense radiation source, that works normally with moderate voltages between 80 to 225 kV. Most X-ray tubes have fixed targets, where the electron beam hits the same spot on the target the whole time. The electron beam current is therefore limited due to heating up this focal spot. For medical X-ray imaging, there are tubes with rotating targets since 1933. The electron beam hits the target not in a single spot, but in a circular path. The load with rotating targets can be enhanced by a factor of approximately ten compared to fixed targets. The reasons why rotating targets are not common in industrial X-ray imaging are locally unstable and quite big focal spots of



(a) Calculated with Lamor formula

(b) Simulated with ROSI

Fig. 8. Analytically calculated and simulated angle distributions for generated X-rays in a LINAC target at high energies

about 800 microns or more and their very high price. They only are used where measurement time is crucial.

With Monte-Carlo simulations some work was done to improve the allowed target load by modifying both the electron beam geometry and target composition with rotating anodes (Sukowski, 2007). This work was done with a medical X-ray tube, but since industrial X-ray tubes are derived from medical tubes, the results can be conveyed to industrial tubes without difficulty. Under variation of the tungsten layer thickness, the emitted X-ray intensity and energy deposition in the target was simulated. The 3-dimensional energy distribution can be transferred to finite element simulation programs to calculate the temperature distribution in steady state while taking cooling effects into account. With the simulation results, optimizing the electron beam and target geometries is possible.

### 3. Simulation of X-ray detectors

#### 3.1 Types of detectors commonly used in NDT

In almost all X-ray imaging applications, line or area pixel sensors are used. An X-ray image is virtually the spatial distribution of the X-ray radiation intensity hitting the sensor area. When X-rays interact with the sensor material, energy is transferred to the sensor and converted into an electrical signal. The signals are amplified and digitized pixel by pixel to a numeric value. The spatial pixel value distribution can be visualized by a color or more often used grey brightness scale. In a positive X-ray image, bright areas correspond to high X-ray intensity, where almost no material is between the X-ray source and the detector, while dark areas are usually covered by thick or heavy parts of the specimen (see 2). In the simulation studies we focused on characterizing flat-panel pixel detectors with squared or rectangular surfaces, which are the most used detectors. Basically there are two types of flat-panel detector technologies that differ in the way of conversion from X-ray energy deposition to an electrical signal (Beutel et al., 2000).

##### 3.1.1 Indirect converting detectors

Most flat-panel detectors convert the X-ray energy deposition in an indirect way into an electrical signal. The X-ray detection mechanism is based on a scintillator. X-rays interacting with a scintillator ionize the atoms, causing emission of fluorescence light due to excited-state

deactivation. The energy level differences of some elements in a typical scintillator are in the range of some electronvolts. The fluorescence light emitted from the scintillator is therefore visual light that can be detected by a photo diode array which is arranged just behind the scintillator layer (Beutel et al., 2000).

### 3.1.2 Direct converting detectors

Unlike scintillator based detectors, direct converting detectors usually consist of a semiconductor material as sensor layer. The semiconductor is assembled between two electrodes. One electrode is continuous over the whole sensor area, while the other electrode on the opposite side consists of many small solder beads which resemble the detector pixels. Between the two electrodes a voltage is applied so that the semiconductor is completely depleted of charge carriers. When X-rays interact with the semiconductor, they transfer energy to bound valence electrons, generating free electron-hole-pairs, which drift to nearby electrode beads due to the electrical field within the semiconductor. At the electrodes a current can be measured, which is proportional to the energy deposited by the X-rays (Beutel et al., 2000).

### 3.1.3 Detector properties

Regardless of application, a perfect detector should fulfill two essential characteristics. First, every X-ray photon hitting the detector surface should create a signal. Since X-rays can pass matter, what makes them useful after all, they also can pass the detector without being detected. The fraction of detected photons  $N_d$  to photons hitting the detector  $N_0$  is not exceeding 1 and is called the detection efficiency  $\eta_{\text{det}}$ .

$$\eta_{\text{det}} = \frac{N_d}{N_0} \leq 1 \quad (4)$$

Especially at high photon energies, the efficiency can be quite low, so the measurement time must be increased for obtaining low-noise images. The efficiency depends on the choice of the sensor material, but mainly on the thickness of the sensor layer. Since X-ray intensity decreases exponentially with the path length in material, increasing the sensor thickness can significantly improve the detection efficiency.

The second important characteristic for spatial resolving detection systems is the ability to determine the location where an X-ray photon hits the detector surface. In the best case, X-rays are not only detected efficiently, they rather should be detected exactly where the initial interaction took place. Unfortunately, this is usually not the case. When X-rays are absorbed by a material, their kinetic energy is transferred to one or more electrons. These electrons propagate through the medium while transferring parts of their kinetic energy to other electrons until stopped. The path length of electrons in matter can reach some tens of microns. Therefore the signal is blurred over a certain volume. Another effect can cause a longer range, but less intense signal blurring. X-rays are not always absorbed by matter, they can also be scattered, transferring only a part of their energy at the location of their initial interaction, what is called Compton scattering. The scattered photon with the remaining energy can be absorbed in a detector volume quite far away (up to some centimeters) from their first point of interaction, causing two or even more signal spots. These two effects occur in both detector types and can be calculated very well by ROSI. They depend on the layer composition (materials and thicknesses) of the detector. In scintillator based detectors there is one more effect that dominates the signal blurring. When X-rays are converted to visual light in the scintillation layer, this light is emitted isotropically to all directions. To be detected, it

has to reach the photo diode layer, where it can be spread over some pixels. This blurring scales highly with the distance from the point of light generation to the photo diode layer. Therefore thick scintillators, where light can be produced quite far away from the photo diode layer often yield a poor spatial resolution. The principle is shown in 9. Signals are clearer distinguishable with thin scintillators, but the efficiency is reduced. Every application demands a different trade-off between efficiency and spatial resolution. The generation, absorption and propagation of visible light in media and on material borders can be described by DETECT2000 (G. McDonald et al., 2000), also a Monte-Carlo simulation code.

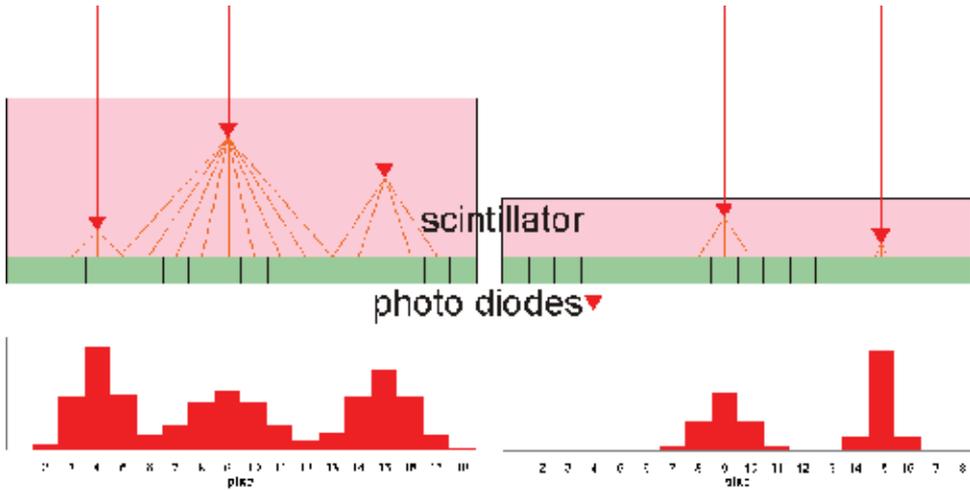


Fig. 9. Signal blurring in a scintillator based detector due to spread of visual photons

For evaluating the relation between the detector properties and their layer composition, one direct converting and one indirect converting detector with 100  $\mu\text{m}$  pixel pitch each were modelled with layer compositions shown in 1.

Detector type	Direct converting (DIC)	Indirect converting (IDC)
Layer composition:		
Front cover	100 $\mu\text{m}$ Al	1 mm Al
Gap	1 mm	1 mm
Front electrode	5 $\mu\text{m}$ Al	
Sensor	750 $\mu\text{m}$ CdTe semiconductor	140 $\mu\text{m}$ Gd <sub>2</sub> O <sub>2</sub> S scintillator
Rear electrodes	50 $\mu\text{m}$ solder	
Fiberoptic plate		3 mm Al <sub>2</sub> O <sub>3</sub>
Electronics	1.5 mm Si	1.5 mm Si
Gap	10 mm	
Rear shielding	2 mm steel	400 $\mu\text{m}$ Cu

Table 1. Detector layer compositions

### 3.2 Simulation of detector properties

#### 3.2.1 Spatial resolution

Like X-ray sources, detectors can be used for a vast amount of applications that demand entirely different properties. Most applications require a good spatial resolution, at least to resolve all details that have to be seen during an inspection. In the last section some effects were introduced that can affect the spatial resolution. The division of the detector in several pixels and their size of course is the most important parameter, but it is mere a numerical issue. The effective spatial resolution can be tested with a double wire test specimen according to the european norm EN462-5. The specimen consists of several pairs of platinum wires with different diameters ranging from 50 to 800 microns. The diameter of each wire of a pair is also the distance between them. This test pattern is placed right in front of the detector entrance window to avoid blurring due to the focal spot. It is also rotated by about 3 degrees to avoid aliasing artifacts. The basic spatial resolution (BSR) can then be derived from the intensity profile perpendicular to the wires. For NDT imaging, the BSR is defined as the theoretical diameter and distance of a wire pair, when the contrast of the space between the wires is at least 20%. To calculate this value, the contrast  $C_{\text{high}}$  of the wire pair with more than 20% contrast (diameter  $d_{\text{high}}$ ) and the contrast  $C_{\text{low}}$  of the wire pair below 20% contrast (diameter  $d_{\text{low}}$ ) is determined. The theoretical diameter  $d_{\text{BSR}}$  of a wire pair with exactly 20% contrast is calculated using linear interpolation.

$$d_{\text{BSR}} = \frac{20\% - C_{\text{low}}}{C_{\text{high}} - C_{\text{low}}} \cdot (d_{\text{high}} - d_{\text{low}}) + d_{\text{low}} \quad (5)$$

The method is also illustrated in 10. The contrast is calculated from the signal differences  $C_{\text{high}} = S_{\text{space,high}}/S_{\text{wire,high}}$  and  $C_{\text{low}} = S_{\text{space,low}}/S_{\text{wire,low}}$ . The BSR is often given as a spatial frequency in line pairs per millimeter.

$$f_{\text{BSR}} = \frac{1}{2 \cdot d_{\text{BSR}}} \quad (6)$$

Another magnitude often used by detector manufacturers is the modulation transfer function (MTF). It is usually measured placing a high absorbing plate in front of the detector with a very sharp and straight edge. The intensity profile perpendicular to the edge is called the edge spread function (ESF), differentiating it results in the line spread function (LSF). The MTF is obtained with fourier transformation of the LSF.

$$MTF(v) = \frac{1}{\sqrt{2\pi}} \int LSF(x) e^{-i2\pi vx} dx = \frac{1}{\sqrt{2\pi}} \int \frac{dESF(x)}{dx} e^{-i2\pi vx} dx \quad (7)$$

The MTF expresses the contrast transfer of a periodic pattern in dependence of the spatial frequency. For ideal pixel detectors without blurring, the MTF becomes a sinc function, where  $p$  is the pixel pitch of the detector.

$$MTF_{\text{ideal}}(v) = \frac{\sin(2p\pi v)}{2p\pi v} = \text{sinc}(2p\pi v) \quad (8)$$

The upper threshold frequency where periodic structures can be reconstructed with any accuracy is called the Nyquist frequency  $\nu_{\text{Nyquist}} = 1/2p$ . As with the BSR, it is assumed that a structure can be resolved at a frequency with at least 20% of contrast transfer.

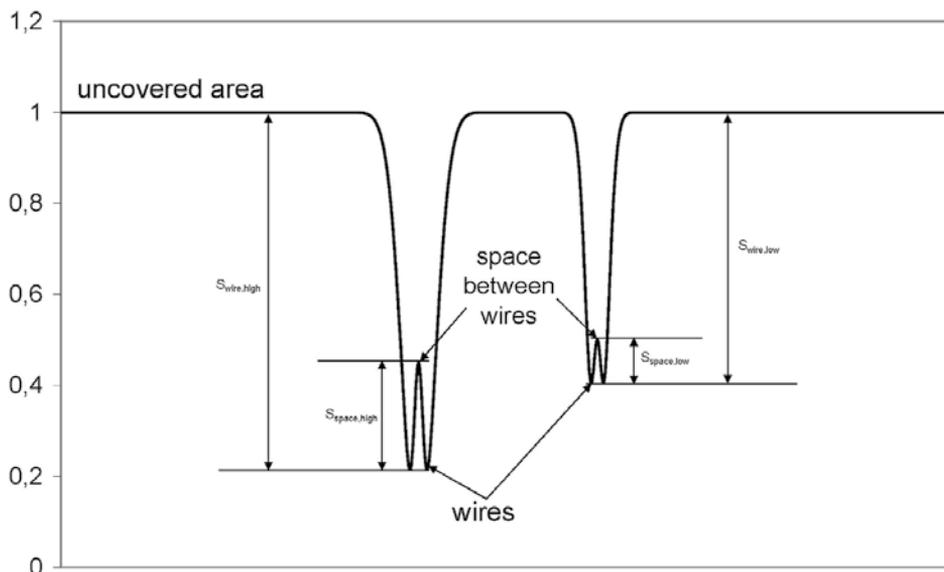


Fig. 10. Method for determining the BSR using the intensity profile along the BSR462-5 double wire specimen

### 3.2.1.1 BSR test results

BSR images were obtained using the following simulation parameters:

- X-ray source: 3 different voltage, prefilter and focal spot size combinations
  - 30 kV, no prefilter, 2  $\mu\text{m}$  focal spot size
  - 160 kV, 4 mm aluminium prefilter, 300  $\mu\text{m}$  focal spot size
  - 450 kV, 4 mm copper prefilter, 2.5 mm focal spot size
- Distance from source to detector: 1 m
- Irradiated detector area: 102.4 mm x 25.6 mm (1024x256 pixels)
- Object placed directly in front of the detector with a rotation of 3 degrees
- Number of simulated photons per image:  $10^9$  ( $\sim 4000$  per pixel)

The images taken with both detectors are shown in 11. The blurring due to optical photon scattering in the image taken with the IDC detector can clearly be seen, the DIC image is quite sharper. The resulting BSR values are shown in 2. In the DIC detector, the signal blurring originates from X-ray photon scattering in the detector volume. Since the scattering cross section increases with photon energy, the BSR values also increase with the mean spectrum energy. In the IDC detector, signal blurring is dominated by scattering of optical photons. The mean interaction depth of photons increases with photon energy, so interactions occur closer to the photo diode matrix. The result is a better resolution with higher energies in contrast to DIC detectors.

### 3.2.1.2 MTF determination

For obtaining MTF images, almost the same parameters were used as for BSR images. To save simulation time, a smaller area of only 12.8 mm x 12.8 mm (128x128 pixels) was irradiated

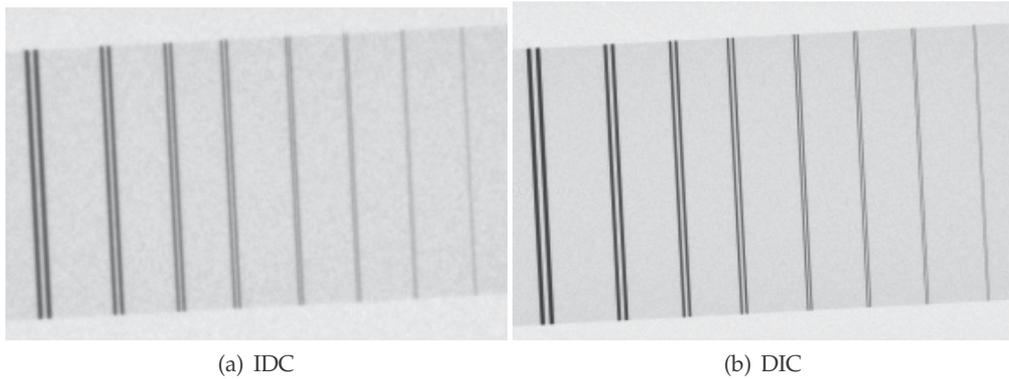


Fig. 11. Images of EN462-5 double wire test pattern

Spectrum	Direct converting (DIC)		Indirect converting (IDC)	
	BSR / $\mu\text{m}$	freq. / lp/mm	BSR / $\mu\text{m}$	freq. / lp/mm
30 kV, no filter	96	5.2	121	4.1
160 kV, 4 mm Al	102	4.9	119	4.2
450 kV, 4 mm Cu	106	4.7	115	4.3

Table 2. BSR values

using  $6.25 \times 10^7$  photons per image. The test object was a 5 mm thick tungsten plate which was also placed directly in front of the detector and rotated by 3 degrees. The edge of the plate leads through the center of the detector (12).

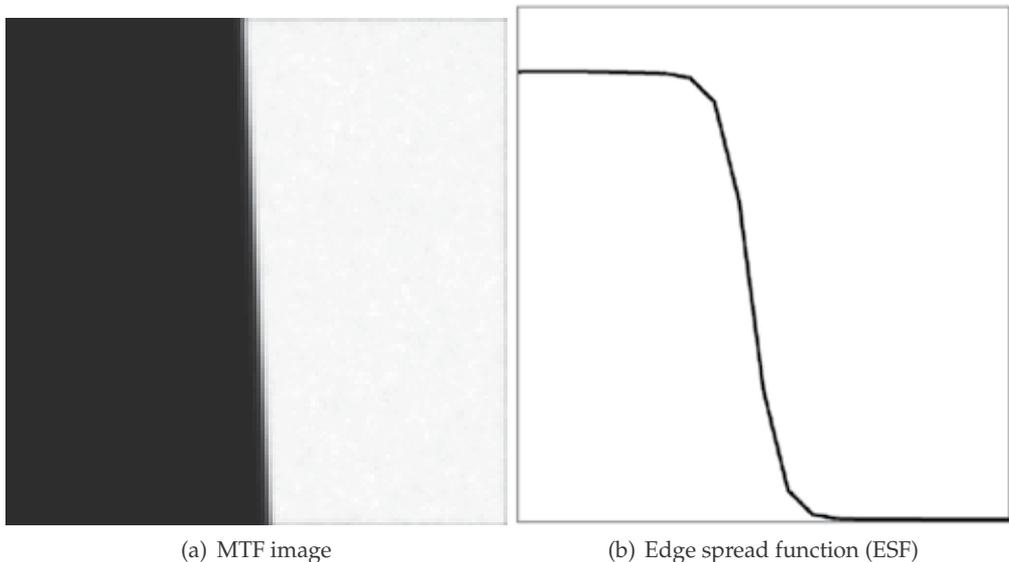


Fig. 12. MTF image and edge spread function taken with IDC detector at 30 kV

Figure 13 shows the calculated MTFs of both detectors. The DIC detector performs better, especially at higher frequencies where optical photon scattering has the largest influence. At

low frequencies on the other hand, the long ranged X-ray scattering processes dominate. The MTF drops quickly at high energies at the beginning of the MTF curve (low frequency drop).

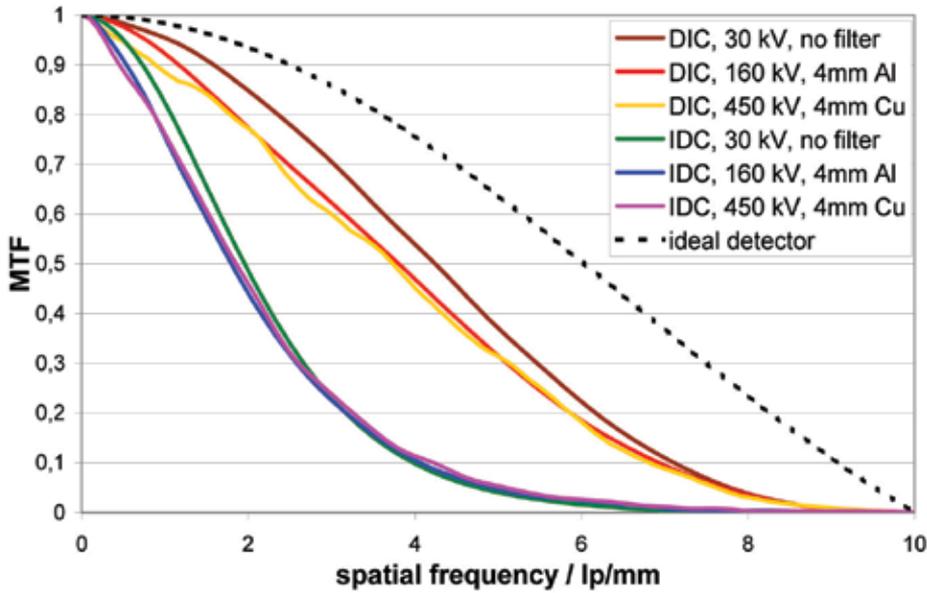


Fig. 13. MTFs for both detectors

**3.2.2 Efficiency**

The efficiency is the increase of the signal-to-noise-ratio (SNR) in a homogenous irradiated image with the radiation dose. The SNR is the mean signal level of the whole image divided by the standard deviation (the noise). The dose is usually measured as an air kerma value, which is the energy deposition in air per air mass. It is measured in Gray (Gy), 1 Gy = 1 J/kg. Table 3 shows the efficiencies of both detectors. The IDC detector shows higher values, since more blurring lowers the pixel variation of signals. At high energies, the DIC detector gets better, because its the sensor layer is very thick (750 μm) compared to that if the IDC detector (140 μm), so high energy photons can still be detected with a fair efficiency.

Spectrum	Direct converting (DIC)	Indirect converting (IDC)
	$SNR / \sqrt{Dose}$ $1 / \sqrt{Gy}$	
30 kV, no filter	3.76	5.09
160 kV, 4 mm Al	29.0	31.6
450 kV, 4 mm Cu	23.5	22.5

Table 3. Efficiency values

**4. Applications**

As already mentioned in the previous section, X-ray Monte-Carlo simulation is a very powerful tool for the design, optimization and the ability to evaluate the proof of concept

of complete X-ray non-destructive-testing (NDT) devices. For X-ray imaging devices e.g. the complete life cycle of each single particle (X-ray photon) including all secondary particles (secondary electrons) can be simulated in detail if needed. The accelerated electrons hitting the tube target emitting bremsstrahlung and characteristic radiation depending on the thickness and layer materials of the target. The generated X-ray photons travel to the specimen and interact via Compton scattering or photoelectric effect. Behind the object the interactions of the photons when hitting the detector can also be studied in detail with all occurring effects like distribution of deposited energy in the detector due to X-ray scattering and the range of the secondary electrons (photo electron).

X-ray system design for the inspection of not yet common specimen, whereupon not yet common means, new in object size, new in material or material combination, new in aspect ratio or also new in the task is sometimes challenging, specially if the specimen and the parameters for X-ray imaging can not be directly derived from former measurements of known objects or the predicted hardware for the inspection system is not available.

Subject to the task inspection systems for non-destructive-testing applications can have different geometries resulting in different requirements for its geometry and used components. A complete overview of X-ray NDT systems and its applications would be go too far but the commonly used principles to mention are radioscopy, computed tomography (CT) and X-ray fluorescence methods.

Independent of the method, the same questions are always of interest when a new inspection system is to be evaluated. Most of interest are boundary conditions like the measurement time or the throughput, the possibility of detection of imperfections or the expected pureness of the separation of the bulk material. The answers are often dependent on each other and the challenge is not only if the task is likely to be solved, but also with what quality at what speed. Therefore derived from the task the system has to be designed in virtual reality and virtual optimizations of the setup have to be done with Monte-Carlo simulations. With the help of simulations the expected performance of the planned system can be predicted.

#### **4.1 Radioscopy**

In radioscopy each specimen is projected on a detector resulting in a 2D image representing the X-ray absorption coefficient of the penetrated material along the X-ray path through the object. With this technique e.g. aluminum casting parts for automotive industry can be inspected and analyzed for defects which might result in a failure of the part during operation. For safety reasons each part in the production line has to be inspected which leads to a need of a very high throughput. The challenge is always to find an optimal trade-off between high throughput and high image quality. The higher the throughput the lower the image quality due to statistical reasons and the lower the performance of the automated image analysis software of the inspection system.

A lot of effects affect the image quality in radioscopy systems. By optimizing the throughput of the inspection system it is of essential interest to suppress all effects reducing the image quality. One effect e.g. is the scattered X-ray radiation from inside the specimen during inspection which hits the detector and reduces the contrast and sharpness of the projection. This effect leads to reduced possibility of detection of small defects. With the Monte-Carlo simulation is it possible to simulate the scattering effects in the specimen and also the distribution of the scattered radiation on the detector. If we know the intensity distribution of the scattered radiation from the specimen on the detector, it can be subtracted from the real image taken during the inspection. With this operation it is possible to get images of the

specimen with nearly no intensity of scattered radiation resulting in better contrast and higher sharpness of the image. In 14 the simulated projection of a step wedge and the simulated intensity distribution of the scattered radiation is shown. Simulation is the only way to get a realistic and not approximated intensity distribution of scattered radiation.

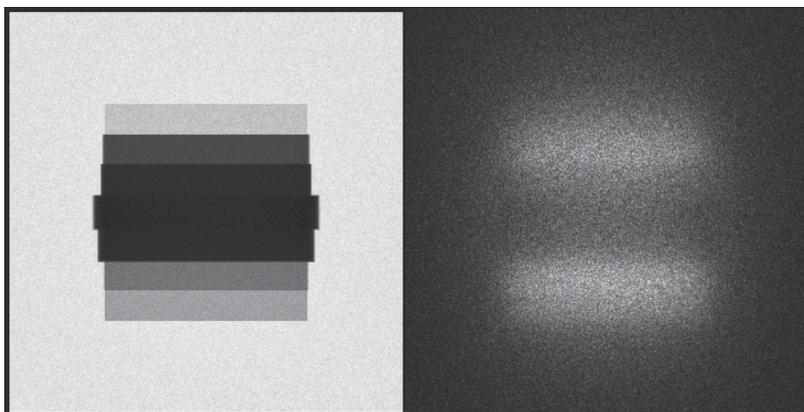


Fig. 14. From left to right: simulated projection of a stepwedge, scattered radiation on the detector.

#### 4.2 Computed tomography (CT)

With computed tomography a 3D distribution of the absorption coefficients of the specimen can be generated providing complete 3D information about the object. The specimen is X-ray projected like many radioscopic images from different angles and the projections can be reconstructed to a 3D volume dataset of the object which can be analysed in 3D.

State of the art e.g. in cargo scanning systems for airport security and customs purposes are 2D scanners providing the personnel only with 2D projections of the freight containers. Due to the overlay projection of different objects in the container the objects often cannot be clearly separated. Driven by this lack of information the idea is to evaluate if it is possible to make a complete CT of the freight container to get the real 3D information. The experimental setup of such a CT system would lead to an investment of expensive equipment. The other way is to virtually design and setup an air cargo scanning system in the Monte-Carlo simulation tool with parameters of real components and make the evaluation with simulations. The virtual setup can be seen in 15.

With this virtual setup in the Monte-Carlo simulation it is possible to predict the expected image quality and recognizability of different materials and objects in an air cargo container together with the scanning times to be expected. In 16 the results of the simulation are shown as reconstructed slices of the air cargo container and its content.

#### 4.3 X-ray fluorescence analysis (XRF)

For the separation of all kinds of bulk material X-ray transmission or X-ray fluorescence methods in combination with a band-conveyor could be a possible solution. Also here is the question at what speed, with what purity and with what spatial resolution the bulk material can be separated. With the Monte-Carlo simulation tool it is possible to simulate the complete process beginning with the optimization of the excitation spectrum, over the excitation of the bulk material with the energy distribution and detection of the excited

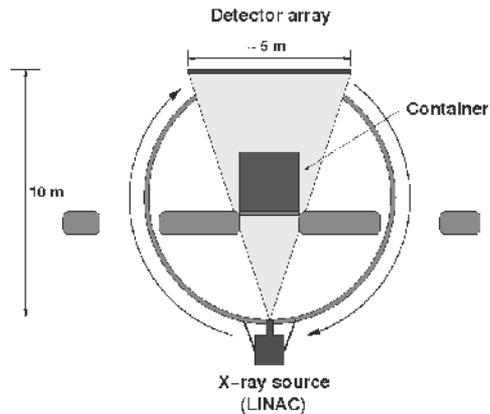


Fig. 15. Sketch of the CT simulation setup of an air cargo container with a LINAC as X-ray source and a 5 m detector array rotating around the container.

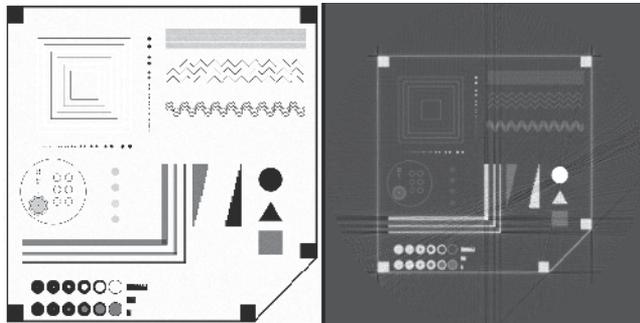


Fig. 16. Left side: One slice of the air cargo container and its content (ideal simulation). Right side: Reconstructed slice of the air cargo container based on simulated projections with reconstruction artifacts due to beam hardening and scattered radiation.

spectrum as a function of different parameters like geometry or X-ray energy. In virtual reality a lot of different parameters in energy, detector systems and geometries can be simulated and evaluated without any real experimental setup. The virtual setup is shown in 17. The simulated detected spectrum of our detector system is shown in 18.

#### 4.4 Dosimetry

Radiation damage due to inspection of some specimen is sometimes a question. For example the dose applied to the content of freight containers or to the electronic parts in PCB inspection systems is of interest to obviate radiation damage and therefore malfunction of the goods. In the MC-Simulation all objects can be defined as detectors which sum up the deposited energy due to the radiation interactions. With summing up the deposited energy it is possible to directly recalculate the applied dose to the specimen in the virtual inspection. With such calculations it is possible to evaluate and predict the applied dose to goods in freight containers which could be expected with a planned inspection system before the system is set up.

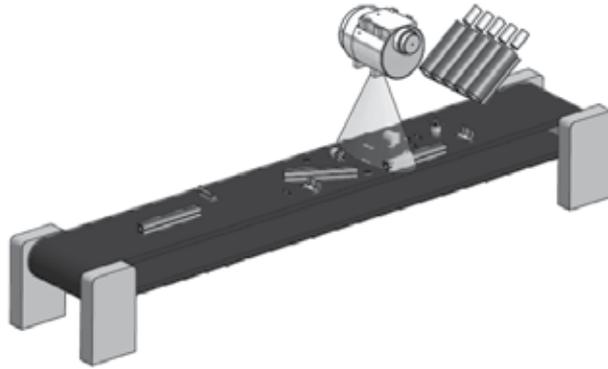


Fig. 17. Setup of the virtual XRF system for evaluation of the expected performance. The high power tube is located above the band-conveyor and to the right of the tube the XRF detector system is located.

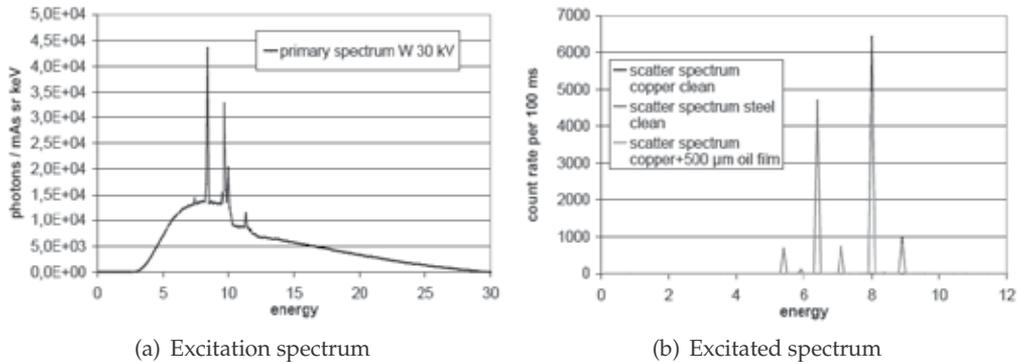


Fig. 18. Simulated excitation spectrum and the resulting excited spectrum of a copper specimen.

## 5. Conclusion

With the X-ray Monte-Carlo simulation ROSI many scenarios can be modelled and calculated realistically. These reach from X-ray generation over imaging applications to X-ray detection. ROSI has also some limitations, since it assumes that electrons and photons have solely particle character. If effects are based on their wave character, another approach has to be done to describe these effects. The simulation of optical light propagation with DETECT2000 is a good example how several simulation codes can be combined to achieve excellent results. Heat generation in X-ray targets and cooling mechanisms can't be described by ROSI directly. But the simulation can provide valuable data for other simulations like finite element programs, where dynamic heat transfer processes can be calculated from 3-dimensional heat energy distributions over the target volume.

Many studies are already done with ROSI to design X-ray targets, detector or whole X-ray devices. The development of ROSI still goes on to include more detailed effects and simulation possibilities.

## 6. References

- Nelson W.R.; Rogers D.W.O. & Hirayama H. (1985). The EGS4 Code System, *Stanford Linear Accelerator Report SLAC-265*, Stanford, CA 94305
- S. Agostinelli et al. (2003). Geant4 - A Simulation toolkit, *Nuclear Instruments and Methods A 506*, pp. 250-303
- H. Morneburg (1995). Bildgebende Systeme für die Medizinische Diagnostik, *SIEMENS Publicis MCD Verlag*, ISBN 978-3895780028
- J. Beutel; H.L. Kundel & R.L. Van Metter (2000). Handbook of Medical Imaging, Volume 1, *SPIE Press*, Bellington, Washington, USA, ISBN 0-8194-3621-6
- J. Giersch & A. Weidemann (2003). ROSI: An object-oriented and parallel computing Monte-Carlo simulation for X-ray imaging, *Nuclear Instruments and Methods A 509*, pp. 151-156
- F. Sukowski (2007). Entwicklung von Hochleistungsröntgenröhren mit Hilfe von Monte-Carlo-Simulationen, *Dissertation*, Friedrich-Alexander-University of Erlangen-Nuremberg, Erlangen
- J.D. Jackson. Klassische Elektrodynamik (2006), *de Gruyter*, ISBN 978-3110189704
- G. McDonald; C. Moisan; F. Cayounet. DETECT2000 the object-oriented version of DETECT, Laval University, Quebec City

# Application of Monte Carlo Simulation in Optical Tweezers

Yu-Xuan Ren<sup>1</sup>, Jian-Guang Wu<sup>2</sup> and Yin-Mei Li<sup>3</sup>

<sup>1,2,3</sup>*University of Science and Technology of China, Hefei, 230026*

<sup>2</sup>*AnHui University of Technology, Maanshan, 243032*

*People's Republic of China*

## 1. Introduction

The concept of optical tweezers(Ashkin et al., 1986) was first conceived by Ashkin et al in 1986. From then on, optical tweezers expands broad research application areas, such as colloidal sciences(Pesce et al., 2009), biophysics(Abbondanzieri et al., 2005; Zhang et al., 2006) and statistical mechanics(Li et al., 2010; McCann et al., 1999). Generally, the probe in optical tweezers' experiment is micrometer-sized or nano- bead, e.g. polystyrene bead, which can be stick to glass surface or chemically linked to biological macromolecules to further reveal the mechanical properties of molecules such as protein or DNA. The trapped bead in optical tweezers is not stationary like mechanical tweezers; it may suffer from random work with its displacement signal obeying Brownian statistics. Analysis of the Brownian motion signal of the trapped probe generates numerous information of the macromolecules, e.g. force, step motion. Therefore, the motion of probe is of great importance in these experiments. Monte Carlo technique provides simulation tools in these experiments to theoretically study the motion of beads in optical tweezers to further reveal the new phenomenon that governs the nature of trapped beads, the characteristics of the optical trap itself and even the mechanical property of macromolecules chemically linked with the trapped microsphere.

The optically trapped microsphere encounters numerous collisions from the surrounding molecules, which constitutes the origin of random forces. The trapped bead behaves like a drunker doing random walk. The topic of analyzing the motion equation of the trapped bead is in the scope of Monte Carlo simulation. In this chapter, we start with the description of light induced radiation force and review the hydrodynamic equation that describes the Brownian motion of trapped bead in optical tweezers in the second and third part, followed by adoption of Monte Carlo simulation in this specific case. In the fourth and fifth parts of this chapter, we show the application methods by presenting two examples of time-sharing optical tweezers and oscillatory optical tweezers in sequence. The sixth part of this chapter discusses potential applications of Monte Carlo simulation in practical colloidal sciences such as artificially induced collision by optical tweezers. This chapter is summarized in the last part.

## 2. Principle of optical tweezers

In macroscopic world, one may use a mechanical tweezers to manipulate an object firmly. What about microscopic manipulation? Optical micromanipulation provides a non-contact,

low destructive and gentle means to manipulate microscopic objects. Ashkin et al had proved that tightly focused laser beam is able to confine microsphere, such as cell, flagellar, colloids etc, due to radiation pressure of light. This concept was then developed to a widely used tool, optical tweezers. The detailed analysis of optical tweezers may be found in many early original literatures and recent review articles. We here briefly review two commonly used methods explaining the forces of a particle in optical tweezers according to the relative geometrical dimension of the particle to that of the wavelength.

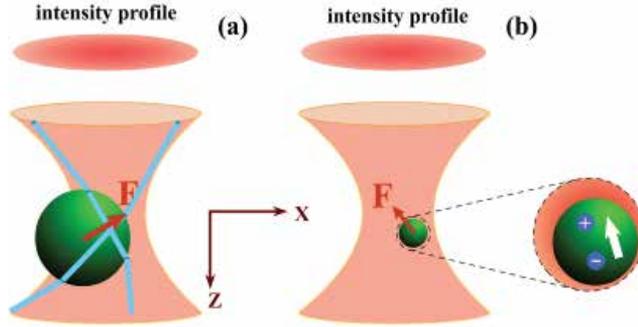


Fig. 1. Principle of optical tweezers. (a) Ray tracing approach based on geometrical optics, (b) multiple dipole model with electromagnetics.

The first approach illustrating the principle of optical tweezers is based on ray optics when the dimension of the particle is greater than the working wavelength as is shown in Fig.1(a). In ray tracing methods, a single optic ray with a portion of power  $P$  is considered hitting the dielectric sphere with incident angle  $\theta$  and incident momentum per second  $n_1P/c$ . The resultant force on the sphere is the sum of contributions due to the reflected ray of power  $PR$  and the infinite number of emergent refracted rays of successively decreasing power  $PT^2, PT^2R, \dots, PT^2R^n, \dots$ . The quantities  $R$  and  $T$  are the Fresnel reflection and transmission coefficients of the surface at  $\theta$ . The net force from the origin can be decomposed into  $F_z$  and  $F_x$  components as follows (Ashkin, 1992)

$$F_z = F_{scat} = \frac{n_1P}{c} \left\{ 1 + R \cos(2\theta) - \frac{T^2 [\cos(2\theta - 2\gamma) + R \cos(2\theta)]}{1 + R^2 + 2R \cos(2\gamma)} \right\} \quad (1)$$

$$F_x = F_{grad} = \frac{n_1P}{c} \left\{ R \sin(2\theta) - \frac{T^2 [\sin(2\theta - 2\gamma) + R \sin(2\theta)]}{1 + R^2 + 2R \cos(2\gamma)} \right\} \quad (2)$$

The total scattering and gradient forces are the sum of the above single ray scattering and gradient forces over all probable incident rays.

For those particles with diameter much less than the working wavelength of laser beam, the scattering force and gradient force are calculated electromagnetically. The dielectric sphere is induced by strongly focused light with multiple dipole, and the resultant force is considered as the interaction between light field and the induced multiple dipole. Followed by (Ashkin et al., 1986), the scattering force is related to the incident power through  $F_{scat} = n_p P_{scat} / c$ , where  $P_{scat}$  is the scattered power. In terms of the intensity  $I_0$  and the effective refractive index  $m$

$$F_{scat} = \frac{I_0}{c} \frac{128\pi^5 a^6}{3\lambda^4} \left( \frac{m^2 - 1}{m^2 + 2} \right)^2 n_p \quad (3)$$

The gradient force  $F_{grad}$  along the direction of the intensity gradient for a Rayleigh spherical particle with polarizability  $\alpha$  is

$$\mathbf{F}_{grad} = -\frac{n_p}{2} \alpha \nabla E^2 = -\frac{n_p^3 a^3}{2} \frac{m^2 - 1}{m^2 + 2} \nabla E^2 \quad (4)$$

Both approaches are able to estimate the trapping ability. We here take electromagnetic model as an example to estimate the trapping behavior. Consider the laser intensity can be rewritten via the time-averaged Poynting vector (Ou-Yang, 1999)

$$I = \langle S \rangle = \frac{1}{2} \frac{1}{\mu_1 n_1 c} |E|^2 = \frac{c \epsilon_0 n_1}{2} |E|^2 \quad (5)$$

The intensity distribution relies on the type of trapping beam used. Many intensity profiles can be used in optical trapping experiments, such as Laguerre-Gaussian beam (Ren, Li, Huang, Wu, Gao, Wang & Li, 2010; Ren, Wu, Zhou, Fu, Sun, Wang & Li, 2010), radially polarized beam (Kozawa & Sato, 2010), etc. Here we consider the most commonly utilized single beam optical trap with Gaussian intensity profile

$$I = I_0 e^{-\frac{r^2}{\omega^2}} \quad (6)$$

where  $\omega$  is the  $1/e$  width of Gaussian intensity profile along the radial direction. Inserting Eqs. 5 and 6 into Eq. 4 yields

$$\mathbf{F}_{grad} = \frac{2n_p^3 a^3}{c \epsilon_0 n} \frac{m^2 - 1}{m^2 + 2} \mathbf{r} e^{-\frac{r^2}{\omega^2}} \equiv k_r \cdot \mathbf{r} e^{-\frac{r^2}{\omega^2}} \quad (7)$$

The gradient force approximately satisfies Hooke's law for significantly small displacements with spring constant  $k_r$ . The spring constant characterizes how stiff the optical trap is, and it has an alternative name, trap stiffness. Experimentally the trap stiffness can be determined by equipartition theorem  $\frac{1}{2} k_B T = \frac{1}{2} k \langle x^2 \rangle$  through measurement of Brownian motion positions, where  $x$  stands for the position of trapped microsphere.

### 3. Dynamics of optically trapped beads

From the microsphere's point of view, it encounters random forces, optical restoring force, frictional force and inertia force. This provides a new approach to evaluate the trap stiffness both experimentally and simulatively. Practically, it is a good approximation utilizing the parabolic potential well model to describe the single beam optical trap. Since the microsphere in the aqueous solution encounters numerous collisions by liquid molecules from all around randomly, the microsphere moves accordingly and the one-dimensional motion equation is characterized by the following Langevin equation

$$m\ddot{x} + \gamma\dot{x} + kx = \sqrt{2k_B T \gamma} \zeta(t) \quad (8)$$

where  $k$  is the static stiffness of an optical trap,  $m$  denotes the mass of the microsphere,  $x(t)$  represents the instantaneous position of the microsphere at time  $t$ ,  $k_B T$  is thermal energy,  $\gamma = 6\pi\eta a$  with  $\eta$  being the viscosity coefficient of surrounding medium, and  $\zeta(t)$  depicts a random Gaussian process satisfying

$$\langle \zeta(t) \rangle = 0 \quad (9)$$

$$\langle \zeta(t)\zeta(t') \rangle = \delta(t - t') \quad (10)$$

In low Reynolds number case, the trapped microsphere is well approximated to an overdamping vibrator, therefore the inertia term is much smaller than the viscous drag force term and can be ignored. The simplified Langevin equation is

$$\gamma\dot{x} + kx = \sqrt{2k_B T \gamma} \zeta(t) \quad (11)$$

Monte Carlo simulation is employed to model the random Gaussian process  $\zeta(t) = \sqrt{-2\ln(u)}\cos(2\pi v)$ , where  $u$  and  $v$  are two uniformly distributed random numbers ranging in  $(0, 1)$ . The simulation algorithm can be deduced from Eq. 8 as follows

$$x_n = x_{n-1} + v_{n-1}\tau \quad (12)$$

$$v_n = v_{n-1} - \frac{kx_{n-1}\tau}{m} + \frac{\sqrt{12\pi k_B T \eta a \tau}}{m} \times \sqrt{-2\ln(u)}\cos(2\pi v) - v_{n-1} \frac{6\pi\eta a}{m}\tau \quad (13)$$

where  $\tau$  indicates the length of the time grid,  $x_n$  is the  $n^{\text{th}}$  elementary position, namely the position at the  $n^{\text{th}}$  time grid,  $v_n$  is the instantaneous velocity of the microsphere correspondingly. Careful attention must be taken to select proper time step to describe the motion of microsphere, since large time step may poorly describe the random process while smaller time step induces longer computation time.

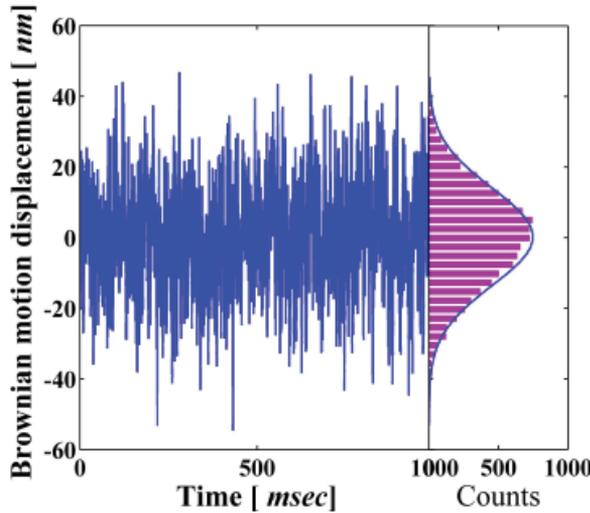


Fig. 2. Brownian motion signal and its histogram. The trapped bead is  $3\mu\text{m}$  diameter polystyrene microsphere and the initial optical trap stiffness  $20\text{pN}/\mu\text{m}$  in the simulation. The histogram indicates the standard deviation is  $\sigma = 13.86\text{nm}$ .

Note that the trap stiffness  $k$  may be numerically calculated *ab initio* through ray optics model or electromagnetic model regarding different sizes of spheres without consideration of optical transmittances and losses of instruments. It is very difficult to predict accurately the actual stiffness if the shape of microsphere is different from each other. Experimentally, due to limited detection speed and systematic noises, the actual stiffness can not be accurately determined either. We here postulate an ideal  $k$  value(true stiffness) throughout Monte Carlo simulation.

The resultant stiffness(measured stiffness) are evaluated through commonly used methods, such as power spectra, equipartition theorem.

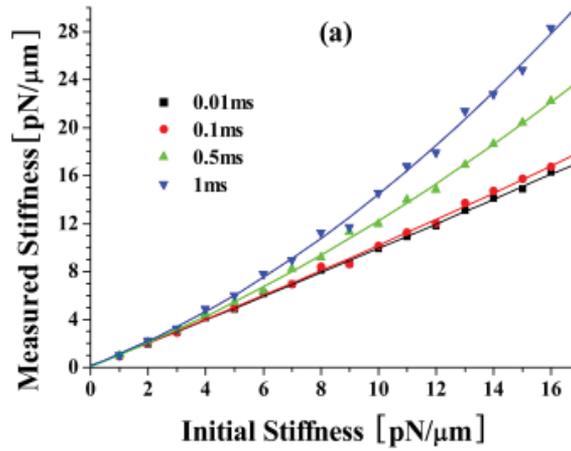


Fig. 3. Measured stiffness varies with respect to initial stiffness under different exposure time performed with Monte Carlo simulation by Gong et al when the trapped bead is  $1\mu m$  diameter polystyrene bead(Gong et al., 2006).

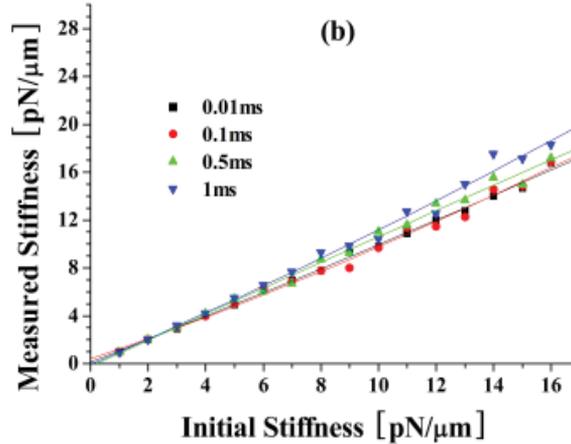


Fig. 4. Measured stiffness varies with respect to initial stiffness under different exposure time performed with Monte Carlo simulation by Gong et al when the trapped bead is  $3\mu m$  diameter polystyrene bead(Gong et al., 2006).

Throughout this chapter, the time step is taken as  $10ns$ . We checked the performance of the algorithm with the following parameters: the initial position of the microsphere is in the trap center with displacement and velocity values all 0. The temperature is  $300K$  with the coefficient of viscosity  $0.801 \times 10^{-3}kg/(m \cdot s)$ . The trapped bead is  $3\mu m$  diameter polystyrene

microsphere with initial optical trap stiffness  $20pN/\mu m$ . The simulated trajectories describing the stochastic motion process of the bead is illustrated in left part of Fig.2. The right hand side of Fig.2 indicates the histogram with its standard deviation  $\sigma = 13.86nm$ . Accordingly, one can calculate the measured stiffness by Monte Carlo simulation through equipartition theorem with the result that  $k_{mea} = k_B T / \langle \sigma^2 \rangle = 21.7pN/\mu m$ , which is greater than the initial stiffness  $20pN/\mu m$ . This indicates that the measured stiffness deviates upwards to the initial stiffness or exact stiffness and is of great significance for actual experiments to select proper experimental parameters, such as exposure time of detector. To further reveal the relationship between the measured stiffness and initial stiffness, Gong et al performed a series of Monte Carlo simulations with different integration times(Gong et al., 2006). Fig.3 shows this relation for  $1\mu m$  diameter polystyrene microsphere with integration time  $0.01ms$ ,  $0.1ms$ ,  $0.5ms$  and  $1ms$  correspondingly denoted by square, circle, upward triangle and downward triangle. The total data collection time in the simulation adopts 5s. We conclude that the higher the acquisition time of the measurement is for a trapping system, the more the measured stiffness values deviate upwards from the initial values.

Another series of simulation were performed for  $3\mu m$  diameter polystyrene spheres with similar conclusions as is illustrated in Fig.4. Two conclusions can be made according to comparison between Figs. 3 and 4. First, for those beads with greater geometrical parameters, the measured stiffness deviates smaller than those with smaller spatial dimension with the same initial stiffness and exposure time. Second, the measured stiffness for less stiffer trap deviates upwards smaller than that for more stiffer trap for polystyrene beads with different sizes. The significance of these findings is that in biophysical experiments, researchers may select significantly larger beads as probe instead of smaller ones when the integration time is not small enough. Alternatively, significantly smaller integration time is employed when the probe adopts smaller nanometer-sized beads, but it cannot be infinitely small due to the limitation of the data acquisition speed of modern detectors.

In order to verify the dependence of measured stiffness with respect to initial stiffness, Gong et al analysed the relation with power spectrum as comparison(Gong et al., 2006). Results from both methods agree well with each other. This reflects that our algorithm performs well and can be used for further research.

#### 4. Effective stiffness of time-sharing optical tweezers

Time-sharing optical tweezers (TSOT) (Liao et al., 2008; Wu et al., 2009) is a very effective technique that produces multiple optical tweezers by splitting a single laser beam at different time intervals to stretch bio-macromolecules or the membrane of human red blood cells. Experimental physicists use various kinds of instruments, such as acousto-optic deflector(AOD)(Emiliani et al., 2004)or a piezoelectric scanning mirror(Mio et al., 2000; Sasaki et al., 1991) to translate light slightly at different frequencies to fulfill quasi-stationary multiple optical traps(Dame et al., 2006).Some novel and practical means are also employed to perform the same function, and a typical example is the rotating glass based time-sharing technique(Ren, Wu, Chen, Li & Li, 2010). Though the above methods employed are different from each other, the realized quasi-stationary traps are all TSOT. In TSOT, The laser beam serves a certain trap at a time interval and immediately switches to another spatial location to serve a new one as is illustrated in Fig.5. For a specific location, the laser switches on and off periodically. Sequential diagram of a trap formed through time-sharing technique is also shown in Fig. 5. The ratios of durations with laser on and off is defined as duty ratio of a TSOT with the following form

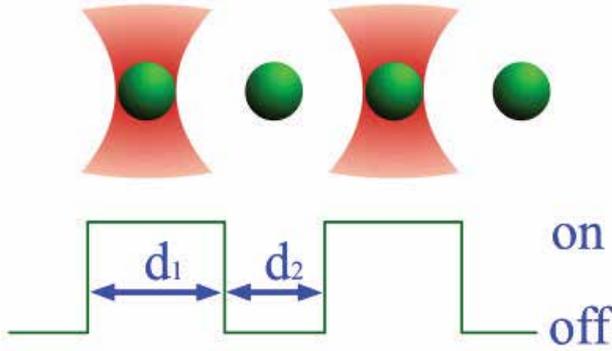


Fig. 5. Schematic diagram of TSOT and its sequential diagram.  $d_1$  and  $d_2$  represent the duration of laser on and off respectively. The periodicity is  $T = d_1 + d_2$ .

$$D = \frac{d_1}{d_2} \quad (14)$$

where  $d_1$  and  $d_2$  represent durations with laser on and laser off respectively in one period, and the sum of them is trap switching periodicity  $T = d_1 + d_2$ . Accordingly, the trap switching frequency  $f_{sw}$  can be written as follows

$$f_{sw} = \frac{1}{d_1 + d_2} \quad (15)$$

Generally, the Brownian motion of a microsphere in aqueous solution is classified into two categories. The first one is that the bead does confined Brownian motion with trapping laser on, while the second is that the bead does free Brownian motion without the exposure of trapping laser. In confined Brownian motion, the trapped bead not only encounters the viscous drag force and stochastic force, but also suffers from restoring force from strongly focused laser beam. The free Brownian motion means that the laser beam does not act on the bead. When the laser switches on one of the multiple trapping positions, the confined Brownian motion of a trapped bead is simulated using Monte Carlo technique. Because the fast beam deflection is very well achieved by the use of AOD or high speed scanning mirror and the rise time to produce different trap positions is of the order of  $\mu s$ , which is smaller than the trap switching periodicity of several orders, the rise time is neglected in our simulation model as a proper assumption.

Due to the limited bandwidth of recent detectors, it is a great challenge to investigate the effective stiffness of TSOT under a broad range of frequency domain. Theoretically, for the motion of microsphere trapped in TSOT, the solution to the Langevin equation can be analytically found by Fourier analysis with a large number of differential equations. Both experiment and theory find it difficult to reveal the variation of effective stiffness with respect to trap switching frequency. Quantitatively, the use of fast beam deflectors is of crucial importance as the time the trap is 'off', servicing another position, has to be much shorter than the time the particle needs to diffuse away from its trapping position (Emiliani et al., 2004). The more time the trap is 'on', the stiffer the trap is. To better understand the stability property of time-sharing multiple optical traps, we use Monte Carlo technique to simulate the motion of a bead in a time-sharing optical trap in a large frequency domain, and numerically calculate the

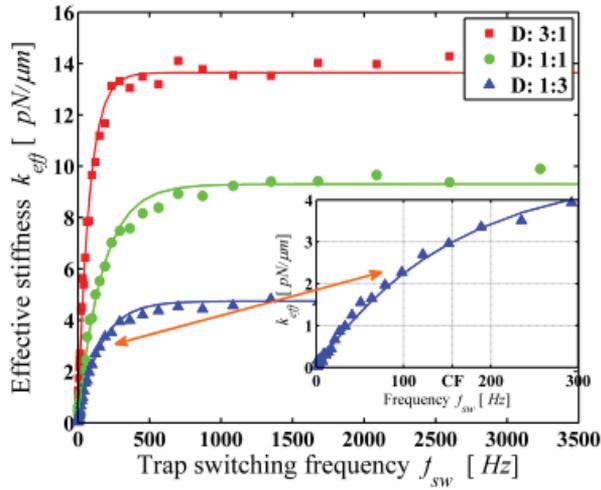


Fig. 6. Effective stiffness varies with respect to trap switching frequency in a long frequency domain for a  $2\mu\text{m}$  diameter polystyrene bead performed by Monte Carlo simulation under TSOT with different duty ratio(Ren, Wu, Zhong & Li, 2010).

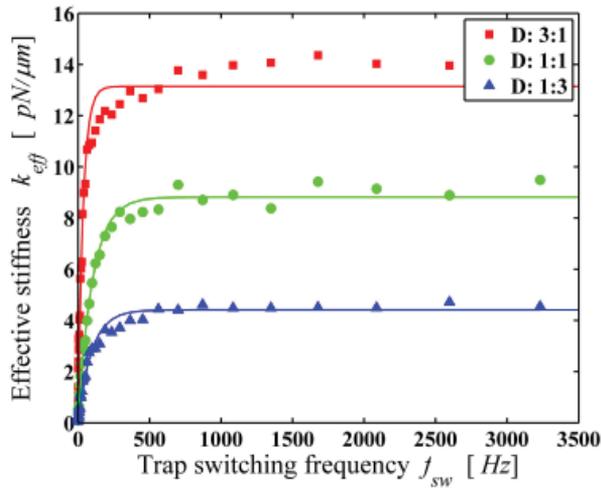


Fig. 7. Effective stiffness varies with respect to trap switching frequency in a long frequency domain for a  $3\mu\text{m}$  diameter polystyrene bead performed by Monte Carlo simulation under TSOT with different duty ratio(Ren, Wu, Zhong & Li, 2010).

effective stiffness of time-sharing optical tweezers according to equipartition theorem. Further the relationship between effective stiffness and trap switching frequency can be revealed. Throughout the simulation, the temperature is set at 298 Kelvin, with drag coefficient of aqueous solution  $0.894 \times 10^{-3} \text{kg}/(\text{m} \cdot \text{s})$ . The microsphere adopts polystyrene bead with mass density being  $1.05 \times 10^{-3} \text{kg}/\text{m}^3$ . Initially, the simulated bead is at the equilibrium position with velocity zero. When the laser is switched on, the stiffness  $k$  in Eq. (3) is equal to  $18 \text{pN}/\mu\text{m}$ , and while the laser off,  $k$  is set to zero during simulation. For continuous wave laser tweezers,

the stability property of a trap is characterized by a stiffness adopting equipartition theory. Similarly, effective stiffness is defined to characterize time-sharing optical tweezers, and it reads

$$k_{eff} = \frac{k_B T}{\sigma^2} \quad (16)$$

In the simulation, the effective stiffness is numerically calculated from 10000 measurements of Brownian motion positions. The simulation shows that for  $2\mu\text{m}$ -diameter polystyrene bead diffused in aqueous solution, the effective stiffness increases with trap switching frequency under different duty ratios as is shown in Fig. 6. The effective stiffness of TSOT with duty ratio 3 : 1 increases sharply with trap switching frequency when the frequency is smaller than 500Hz; while the frequency is larger than 500Hz, the effective stiffness does not vary with respect to trap switching frequency. This general relation holds for other duty ratios, e.g.  $D = 1 : 1$  and  $D = 1 : 3$ . The relation is linear when the frequency domain is significantly small for TSOT with duty ratio 1 : 3 as illustrated in inset of Fig. 6. In low frequency range, this relation is linear and is verified experimentally by Ren et al through glass plate based TSOT in the frequency range from 5Hz to 70Hz(Ren, Wu, Chen, Li & Li, 2010). Similar results for  $3\mu\text{m}$ -diameter polystyrene bead are shown in Fig. 7 from Monte Carlo simulation. The results for beads with diameter of both  $2\mu\text{m}$  and  $3\mu\text{m}$  indicate a general trend for the relationship between effective stiffness and switching frequency.

To analytically describe the dependence of stiffness on the trap switching frequency, we employ Box Lucas Model(Box & Lucas, 1959), which was first introduced to describe the quantum yield of intermediate product of a consecutive chemical reaction, to fit our simulation dependence of effective stiffness on trap switching frequency. The analytical relation is qualitatively described by  $k_{eff} = k_0 \cdot (1 - \exp(-f_{sw}/f_{ch}))$ , where  $k_0$  and  $f_{ch}$  stand for transient-free stiffness and characteristic frequency respectively. For  $2\mu\text{m}$  diameter polystyrene bead trapped in TSOT with duty ratio 1 : 3, the fitted values of the two parameters are:  $k_0 = (4.74 \pm 0.05)(\text{pN}/\mu\text{m})$  and  $f_{ch} = (156 \pm 6)(\text{Hz})$  with coefficient of determination  $R^2$  being 0.9839 which means the Box Lucas Model well describes the dependence of effective stiffness on switching frequency in a large frequency domain. In inset of Fig.6 there are two letters CF marked on the horizontal axis, and this mark indicates where characteristic frequency locates for  $2\mu\text{m}$ -diameter bead trapped in TSOT with duty ratio 1 : 3. Following the same procedure, the parameters  $k_0$  and  $f_{ch}$  for beads with different diameters in TSOT with different duty ratios are summarized in Table 1.

Table 1. Summation of  $k_0$  and  $f_{ch}$  of time-sharing optical tweezers with different duty ratios(Ren, Wu, Zhong & Li, 2010).

Duty ratio	$2\mu\text{m}$		$3\mu\text{m}$	
	$k_0$ (pN/ $\mu\text{m}$ )	$f_{ch}$ (Hz)	$k_0$ (pN/ $\mu\text{m}$ )	$f_{ch}$ (Hz)
3 : 1	$13.65 \pm 0.12$	$78 \pm 3$	$13.14 \pm 0.19$	$36 \pm 2$
1 : 1	$9.30 \pm 0.10$	$169 \pm 7$	$8.82 \pm 0.09$	$102 \pm 4$
1 : 3	$4.74 \pm 0.05$	$156 \pm 6$	$4.41 \pm 0.06$	$95 \pm 5$

Actually, in higher frequency ranges, such as the case of femtosecond laser optical tweezers(Zhou et al., 2008), the effective stiffness doesn't vary with the increase of modulation frequency. The trap is stable subjected to change of repetition rate of femtosecond laser with a fixed output power, and the effective stiffness is only affected by the average output power of laser with high repetition rate.

In lower frequency range, the model can be approximated by a linear regression model at high accuracy. When the switching frequency is smaller than the characteristic frequency, namely  $f_{sw} < f_{ch}$ , simulation results verify that it performs well even using linear regression model, which is of great importance when using unstable TSOT with low switching frequencies such as the case of studying the colloidal collision frequency.

The simulation results indicate that for a certain bead trapped in TSOT with different duty ratios the transient-free stiffness  $k_0$  increases with duty ratio which determines the average power of a certain trap. As for the same bead, the characteristic frequency varies with the duty ratio, and according to our simulation, the characteristic frequency with duty ratio 3:1 is smaller than those with other two duty ratios both for beads with diameter  $2\mu m$  and  $3\mu m$ . A proper explanation is that the effective stiffness transits to a stable value quicker than that with small duty ratio when increasing the trap switching frequency. Meanwhile, the characteristic frequencies with duty ratios 1:1 and 1:3 are larger and close to each other.

## 5. Dynamics of microsphere under oscillatory optical tweezers

Oscillatory optical tweezers is another derivative of spatially and temporally modulated optical trap. Oscillatory optical tweezers plays its unique role in microrheology and biophysics. It can be used to measure the elastic storage modulus and viscous loss modulus of a material, elasticity of DNA or red blood cell, etc. Oscillatory optical tweezers can be constituted by moving the center of a single beam optical trap sinusoidally with a fixed frequency  $\omega$  as is shown in Fig.8.

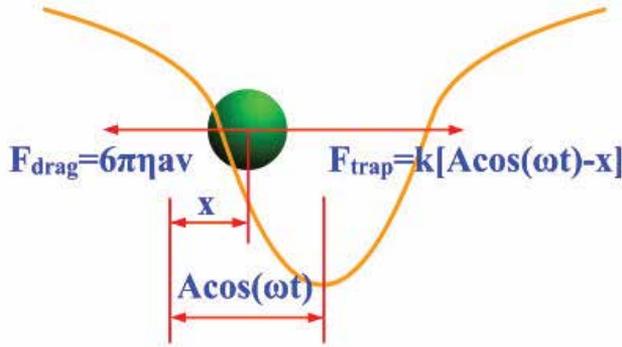


Fig. 8. Schematic diagram representing oscillatory optical tweezers.

Therefore substituting  $kx$  in Eq.8 by  $k[x - A \cos(\omega t)]$  yields the Langevin equation in oscillatory optical tweezers case, and it reads

$$m\ddot{x} + \gamma\dot{x} + k[x - A \cos(\omega t)] = \sqrt{2k_B T \gamma} \zeta(t) \quad (17)$$

where  $A$  stands for the amplitude of oscillation of trap center,  $\omega$  is the oscillation frequency. The stiffness of a single beam optical tweezers  $k$  is not a constant value when the displacement is large enough from the trap center as is shown in Eq.7 and varies nonlinearly with displacement when the position is far away from the center. In this work, we assume the colloidal particle is trapped in the linear region of optical tweezers with  $k$  being constant value and the amplitude  $A$  is much smaller than the radius of linear region. Similar to the algorithms described in Eq. 12 and 13, the Monte Carlo simulation algorithms can be written as following in the oscillatory optical tweezers' case

$$x_n = x_{n-1} + v_{n-1}\tau \quad (18)$$

$$v_n = v_{n-1} - \frac{k[x_{n-1} - A\cos(\omega n\tau)]\tau}{m} + \frac{\sqrt{12\pi k_B T \eta a \tau}}{m} \times \sqrt{-2\ln(u)\cos(2\pi\nu)} - v_{n-1} \frac{6\pi\eta a}{m} \tau \quad (19)$$

In the simulation, amplitude  $A$ , time step  $\tau$ , radius of polystyrene bead adopt  $50nm$ ,  $1ns$ ,  $100nm$  correspondingly. The coefficient of viscosity is  $0.80 \times 10^{-3}kg/(m \cdot s)$  at the temperature of  $303K$ . The trap stiffness adopts  $8pN/\mu m$ . Monte Carlo simulation procedures are performed with different oscillation frequencies to generate sequential displacements of trapped sphere.

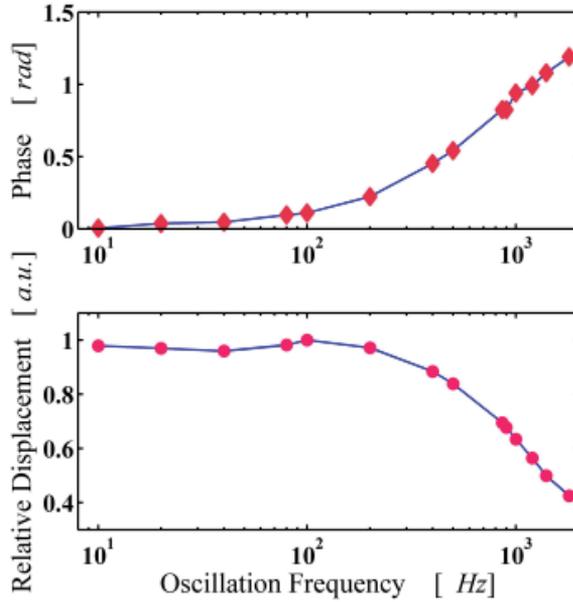


Fig. 9. Phase and displacement of colloidal particles vary with frequencies in oscillatory optical tweezers.

In the simulation, the dynamics of colloidal particle are studied from the output states, which were deduced from the simulated input states (Seol et al., 2004) of colloidal particle encountering in oscillatory optical tweezers. The input states are simulated utilizing the differential form of Eqs. 18 and 19 similar to the procedure of time-sharing optical tweezers (Ren, Wu, Zhong & Li, 2010) or jumping optical tweezers (Liao et al., 2008). Digital lock-in amplifier is adopted to further analyze the simulated displacement signals. The relationship between the movement of the bead and the oscillation of the trap center can be clearly seen by this procedure as is shown in Fig.9. The phase retardation increases with frequency while the amplitude decreases with frequency.

Additionally, when there exists external sources of noise contributing to the motion of the particle, power spectrum analysis was chosen to clearly see the external noise (Horst & Forde, 2008), such as the oscillation of the trap center in oscillation optical tweezers case. In order to analyze the influence of oscillation of the trap center, power spectrum of Brownian positions of colloidal particle with diameter  $200nm$  in oscillatory optical tweezers is analyzed as illustrated in Fig. 10. The curves demonstrated in Fig. 10 show that there exists a resonant peak

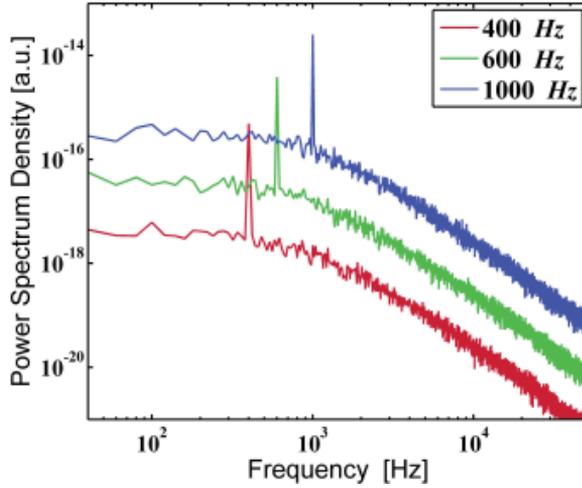


Fig. 10. Power spectrum of colloidal particles trapped in oscillatory optical tweezers with different oscillation frequencies.

typically around 400Hz, 600Hz and 1000Hz correspondingly. These resonant frequencies are in accordance with the oscillation frequencies adopted in the simulation. In our simulation, the corner frequency is  $f_c = k / (2\pi\gamma) = k / (12\pi^2\eta a) = 8 / (12\pi^2 \times 0.8 \times 10^{-3} \times 0.1 \times 10^{-6}) \text{Hz} = 844 \text{Hz}$  in the system. The observed peak may be explained by resonance of the particle system to external oscillation.

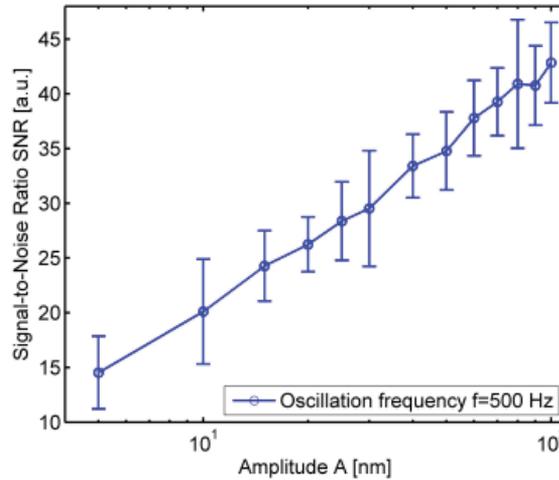


Fig. 11. Signal noise ratio as a function of amplitude of oscillation.

In order to qualitatively investigate the intensity of resonant peak with respect to the oscillation amplitude of the trap center, the signal-noise ratio  $SNR$  was introduced to characterize the resonance,

$$SNR = 10 \log \frac{S(\omega)}{S'(\omega)} \quad (20)$$

where  $S'(\omega)$  represents the power spectrum value at the resonant frequency  $\omega$  without oscillation of the trap center and is evaluated from the power spectrum values around resonance peak, and  $S(\omega)$  represents the power spectrum of colloidal particle trapped in oscillatory optical tweezers at frequency  $\omega$ . Another series of simulations were performed with oscillation frequency 500Hz and the microsphere adopts 200nm diameter polystyrene bead.

Simulation results demonstrate that the SNR is proportional to the logarithm of oscillation amplitude  $A$  of the trap center. This is of great significance to excite oscillation with proper SNR by varying the oscillation amplitude and will have potential applications in measurement of trap stiffness, drag coefficient, particle size and even temperature of the surrounding medium because the resonance intensity is closely related with these parameters.

## 6. Conclusions

This chapter presents primary examples using Monte Carlo technique in optical tweezers related research and applications. Since optical tweezers has broad applications in colloidal sciences, biophysics, nanotechnology etc., Monte Carlo simulation will find its applications in these interdisciplinary researches. At list 4 topics can be studied indeep, (1) translational diffusion behaviors of colloids under different solvent conditions; (2) rotational diffusion properties induced by strongly focused light with orbital angular momentum; (3) rotational Brownian motion under optical torque for birefringence of innate rotating cell such as motor flagellar; (4) collision of microspheres induced by optical tweezers revealing the stability of colloids.

## 7. References

- Abbondanzieri, E. A., Greenleaf, W. J., Shaevitz, J. W., Landick, R. & Block, S. M. (2005). Direct observation of base-pair stepping by RNA polymerase, *Nature* 438(7067): 460–465.
- Ashkin, A. (1992). Forces of a single-beam gradient laser trap on a dielectric sphere in the ray optics regime, *Biophys. J.* 61: 569–582.
- Ashkin, A., Dziedzic, J. M., Bjorkholm, J. E. & Chu, S. (1986). Observation of a single-beam gradient force optical trap for dielectric particles, *Opt. Lett.* 11(5): 288–290.
- Box, G. E. P. & Lucas, H. L. (1959). Design of experiments in non-linear situations, *Biometrika* 46: 77–90.
- Dame, R. T., Noom, M. C. & Wuite, G. J. L. (2006). Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation, *Nature* 444(16): 387–390.
- Emiliani, V., Sanvitto, D., Zahid, M., Gerbal, F. & Coppey-Moisan, M. (2004). Multi force optical tweezers to generate gradients of forces, *Opt. Express* 12(17): 3906–3910.
- Gong, Z., Chen, H.-T., Xu, S.-H., Li, Y.-M. & Lou, L.-R. (2006). Monte-carlo simulation of optical trap stiffness measurement, *Opt. Commun.* 263: 229–234.
- Horst, A. v. d. & Forde, N. R. (2008). Calibration of dynamic holographic optical tweezers for force measurements on biomaterials, *Opt. Express* 16(25): 20987–21003.
- Kozawa, Y. & Sato, S. (2010). Optical trapping of micrometer-sized dielectric particles by cylindrical vector beams, *Opt. Express* 18(10): 10828–10833.
- Li, T., Kheifets, S., Medellin, D. & Raizen, M. G. (2010). Measurement of the instantaneous velocity of a Brownian particle, *Science* 328(5986): 1673–1675.

- Liao, G.-B., Bareil, P. B., Sheng, Y. & Chiou, A. (2008). One-dimensional jumping optical tweezers for optical stretching of bi-concave human red blood cells, *Opt. Express* 16(3): 1996–2004.
- McCann, L. I., Dykman, M. & Golding, B. (1999). Thermally activated transitions in a bistable three-dimensional optical trap, *Nature* 402: 785–787.
- Mio, C., Gong, T., Terray, A. & Marr, D. W. M. (2000). Design of a scanning laser optical trap for multiparticle manipulation, *Rev. Sci. Instrum.* 71(5): 2196–2200.
- Ou-Yang, H. D. (1999). Design and applications of oscillating optical tweezers for direct measurements of colloidal forces, *Colloid-Polymer Interactions: From Fundamentals to Practice*, John Wiley and Sons, New York.
- Pesce, G., Luca, A. C. D., Rusciano, G., Netti, P. A., Fusco, S. & Sasso, A. (2009). Microrheology of complex fluids using optical tweezers: a comparison with macrorheological measurements, *J. Opt. A: Pure Appl. Opt.* 11: 034016.
- Ren, Y.-X., Li, M., Huang, K., Wu, J.-G., Gao, H.-F., Wang, Z.-Q. & Li, Y.-M. (2010). Experimental generation of Laguerre-Gaussian beam using digital micromirror device, *Appl. Opt.* 49(10): 1838–1844.
- Ren, Y.-X., Wu, J.-G., Chen, M., Li, H. & Li, Y.-M. (2010). Stability of novel time-sharing dual optical tweezers using a rotating tilt glass plate, *Chinese Physics Letters* 27(2): 028703.
- Ren, Y.-X., Wu, J.-G., Zhong, M.-C. & Li, Y.-M. (2010). Monte carlo simulation of effective stiffness of time-sharing optical tweezers, *Chinese Optics Letters* 8(2): 170–172.
- Ren, Y.-X., Wu, J.-G., Zhou, X.-W., Fu, S.-J., Sun, Q., Wang, Z.-Q. & Li, Y.-M. (2010). Experimental generation of Laguerre-Gaussian beam using angular diffraction of binary phase plate, *Acta Physica Sinica(In Chinese)* 59(6): 3930–3935.
- Sasaki, K., Koshioka, M., Misawa, H., Kitamura, N. & Masuhara, H. (1991). Pattern formation and flow control of fine particles by laser-scanning micromanipulation, *Opt. Lett.* 16(9): 1463–1465.
- Seol, Y., Visscher, K. & Walton, D. B. (2004). Suppression of noise in a noisy optical trap, *Physical Review Letters* 93(16): 160602.
- Wu, J.-G., Ren, Y.-X., Wang, Z.-Q., Zhou, C. & Li, Y.-M. (2009). Time-sharing multiple optical tweezers using rotating glass plate, *Chinese J. Lasers (In Chinese)* 36(10): 2751–2756.
- Zhang, Y., Smith, C. L., Saha, A., Grill, S. W., Mihardja, S., Smith, S. B., Cairns, B. R., Peterson, C. L. & Bustamante, C. (2006). DNA translocation and loop formation mechanism of chromatin remodeling by SWI/SNF and RSC, *Molecular Cell* 24(4): 559–568.
- Zhou, M., Yang, H., Di, J. & Zhao, E. (2008). Manipulation on human red blood cells with femtosecond optical tweezers, *Chin. Opt. Lett.* 6(12): 919–921.

# Enabling Grids for GATE Monte-Carlo Radiation Therapy Simulations with the GATE-Lab

Sorina Camarasu-Pop<sup>1</sup>, Tristan Glatard<sup>1</sup>, Hugues Benoit-Cattin<sup>1</sup>  
and David Sarrut<sup>2</sup>

<sup>1</sup>*Université de Lyon, CREATIS; CNRS UMR5220; Inserm U1044;  
INSA-Lyon; Université Lyon 1*

<sup>2</sup>*Université de Lyon, CREATIS; CNRS UMR5220; Inserm U1044;  
INSA-Lyon; Université Lyon 1; Centre Léon Bérard  
France*

## 1. Introduction

Among radiation therapy simulation methods, Monte-Carlo approaches are known to be the most accurate but they are heavy to use because of their computing time. Nowadays they can be accelerated with the help of the ever-increasing computing power and distributed resources mutualised in clusters, clouds or grids (Montagnat et al. (2005)).

Grid infrastructures are used both for experimental and production purposes. They have been designed to support data and computing requirements for a large spectrum of applications from various scientific domains. Nevertheless they are not yet at a "plug and play" phase that would allow applications to be easily and efficiently deployed on the existing infrastructure. Efforts are required to achieve reliable and efficient execution on the grid for new applications and to provide user-friendly execution environments to end-users.

This chapter presents a solution for reliable, user-friendly and fast execution of GATE (Jan et al. (2004)) on a grid. Developed within the OpenGate<sup>1</sup> international collaboration, GATE is a Monte-Carlo based open-source software for nuclear medicine simulations, especially for TEP and SPECT imaging, and for radiation therapy applications. The solution proposed here enables transparent grid execution from a user-friendly interface. The application is parallelized automatically and a dynamic partitioning can also be used for further reducing the execution time and improving the robustness to job failures (Camarasu-Pop et al. (2010)). This chapter will discuss in more detail the implemented solution by describing the user interface, the system architecture, as well as the dynamic optimization strategy. Usage and performance results illustrating the system adoption will then be presented. To conclude with, it will discuss the lessons learned in building the system.

## 2. Related work

Different parallelization methods for Monte-Carlo simulations have been proposed for execution on distributed environments. This section will present related work on

---

<sup>1</sup> <http://opengatecollaboration.healthgrid.org/>

parallelization methods, on challenges raised by distributed environments (e.g. production grids), as well as on end user interfaces allowing to run such parallelized applications on grids.

### 2.1 Parallelization methods

The simulation in a particle tracking Monte-Carlo system consists in the successive stochastic tracking through matter of a large number of individual particles. Each particle has an initial set of properties (type, location, direction, energy, etc) and its interaction with matter is determined according to realistic interaction probabilities and angular distributions. Accurate results require the simulation of a large number of particles and physical interactions can also produce other particles that must be tracked. Therefore, typical radiation therapy simulations can take several days or even weeks to complete on a single computer. However, they can be easily parallelized on distributed systems. Instead of sequentially simulating a large number (up to several billions) of particles, smaller groups (bags) of particles can be simulated independently and eventually merged. This process is valid only if the sub-simulations are statistically independent, which in GATE is guaranteed by using a special random number generator as explained by Reuillon et al. (2008).

Among existing parallelization methods, the simplest and most commonly used is particle parallelism. In this case the geometry information is replicated on each processor and particles are distributed between available processors as described in Maigne et al. (2004). The authors present first results obtained by running GATE in parallel on multiple processors of the DataGrid<sup>2</sup> project. In this example the number of particles simulated on each of the  $N$  processors represents a fraction  $P/N$  of the total of  $P$  particles. This static distribution of particles often underexploits resources when the simulation is executed on heterogeneous platforms like computer grids. Indeed, as particles are evenly distributed among tasks, tasks running on fast resources complete before other ones and these resources may remain idle if the scheduler cannot assign them other tasks (e.g. at the end of the simulation). Moreover, it may happen that a task is allocated to a slow resource towards the end of the simulation, thus slowing down the completion of the whole application (Cirne et al. (2007)).

A possible solution to this problem is the dynamic distribution and/or reassignment of particles to available processors during runtime. Dynamic partitioning is proposed in Galyuk et al. (2002) and in Procassini et al. (2005) for spatial parallelism. Spatial parallelism involves splitting the geometry into domains and then assigning a specific domain to one processor. This method is usually needed when the problem geometry has a significant size so that one processor does not have enough memory to store all particles/zones. Spatial parallelism may introduce load imbalance between processors, as spatial domains will require different amounts of computational work. In Procassini et al. (2005), a dynamic load balancing algorithm distributes the available processors to the spatial domains. Communications are generated between processors to transmit changes from the last state. In Galyuk et al. (2002), the parallelization is done using an MPI implementation and is based on a semaphore principle under distributed memory conditions. These two implementations are therefore cluster oriented and are not adapted for grid usage where communications between processors are very costly.

Camarasu-Pop et al. (2010) propose a dynamic partitioning algorithm for GATE radiotherapy simulations using pilot jobs. Statistically independent simulations are launched on available resources and they keep on running until the desired number of particles is reached.

---

<sup>2</sup> <http://eu-datagrid.web.cern.ch>

The simulation is no longer split into sub-simulations; instead each computing resource contributes to the whole simulation until it is completed. The pilot-job master periodically sums up the number of simulated particles and sends stop signals to tasks when needed. Due to the randomness involved in Monte-Carlo simulations, this is possible with no communication between tasks.

## 2.2 Applications execution on grids

Significant work has been put in parallelizing Monte-Carlo algorithms and running them on distributed systems. Nevertheless, they are rarely accessible to physicians and researchers because of the recurrent errors and complexity introduced by grid infrastructures. Grid middleware has to be coupled with other tools to ensure quality of service (QoS) (Tan et al. (2010)), to facilitate application porting (Camarasu-Pop et al. (2008)) and to offer high-level execution interfaces (Olabarriaga et al. (2010)).

Failures are recurrent on large grid infrastructures like the EGI grid, where the success rate is usually of 80 to 85% (Jacq et al. (2008)). When splitting one Monte-Carlo simulation into sub-tasks, it is important that all of them complete successfully in order to retrieve the final result. Therefore, failed tasks must be resubmitted, further slowing down the application completion. Frameworks such as pilot jobs have been introduced during the last years (Ahn et al. (2008); Bagnasco et al. (2008); Kacsuk et al. (2008); Maeno (2008); Moscicki (2003b); Sfiligoi (2008); Tsaregorodtsev et al. (2008)) and are now extensively used in order to improve QoS and cope with the recurrent errors and high latencies caused by the grid heterogeneity.

Application porting and reusability have also been facilitated with the help of workflow technology, as discussed in Deelman et al. (2005); Glatard et al. (2008); Kacsuk & Sipos (2005); Maheshwari et al. (2009); Oinn et al. (2004). Engines now allow the execution of workflow applications on various grid middleware in a generic way.

On top of these tools, high-level execution interfaces are essential for end users with no specific grid knowledge. This problem has been acknowledged by different communities who already developed a certain number of grid portals such as Genius<sup>3</sup> or GridSphere<sup>4</sup>. A few communities have customized these tools for their needs or even developed their own specific portals. This is also the case for some of the Monte-Carlo applications running on clusters or grids as detailed hereafter.

## 2.3 End user interfaces/portals for Monte-Carlo applications

Early portals for Monte-Carlo applications were designed to provide physicists convenient access to grid tools and services, as presented by Engh et al. (2003) and Compostella et al. (2007). These portals provided functionalities like job submission and monitoring, as well as results browsing. Nevertheless, they were setup at a time when grid reliability was poor and end users still had to handle significant debugging processes.

Recent portals integrate increasingly powerful functionalities and are often targeting specialized communities.

The E-IMRT platform described in José Carlos Mouriño Gallego (2007); Pena et al. (2009) offers radiotherapists a set of algorithms to optimize and validate radiotherapy treatments. It has three main components, namely characterization of linear accelerators, radiotherapy treatment planning optimization and verification and data repository. The IMRT (Intensity-Modulated Radiation Therapy) treatment verification is based on a Monte-Carlo

<sup>3</sup> <https://genius.ct.infn.it/>

<sup>4</sup> <http://www.gridisphere.org>

method and as such is an excellent candidate to be executed on a grid. These services are accessible through an user-friendly web page, where the implementation of the verification and optimization algorithms, as well as the complexity of the distributed infrastructure on which they run, are hidden to the user. The E-IMRT platform became one of the 25 Grid Business experiments (BEs) from the BEinGRID (Business Experiments in Grid) project. BEinEIMRT (M.G. Bugeiro (2009)) provides on-demand e-Health computational services (like tumor detection and radiotherapy planning) to Health organisations like clinics and hospitals. All services are provided by the Centro de Supercomputación de Galicia (CESGA), which is the customary provider. When CESGA is under peak demand, external resources are added for a limited time period<sup>5</sup>. This process is transparent for the end user, reducing execution time and increasing the QoS.

The HOPE (HOspital Platform for E-health) platform<sup>6</sup> was designed to enable GATE simulations in a grid environment. Its user-friendly interface was built taking into account feedback from healthcare professionals. Like the E-IMRT platform, it targeted especially physicians and medical physicists.

The PARTNER (PARTicle Training Network for European Radiotherapy) project aims at reinforcing research and training professionals in the rapidly emerging field of hadron therapy. The project started at the end of 2008 for a period of 4 years. It has reviewed<sup>7</sup> existing ICT (Information and Communications Technology) based medical collaborative infrastructures and by the end of the project it will build a grid testbed allowing the training of a new generation of researchers.

Researchers' requirements are different from those of the clinicians and medical physicists who are targeted by the E-IMRT and HOPE platforms. These platforms were mainly designed to perform the same (limited) set of (standard/validated) applications to several data: for example computing IMRT treatment plans on different patient data. In contrast, researchers need to be able to easily test simulations with different sets of parameters, to perform computing intensive simulations that lead to large phase-spaces or to study the design of new imaging devices, such the ones that will be developed to monitor the dose deposit in hadrontherapy situations thanks to the outgoing particles produced by nuclear interactions.

### 3. The GATE-Lab

The GATE-Lab targets mainly researchers, both in academic and R&D industrial environments. Indeed, GATE is designed for a wide range of purposes in the fields of PET, SPECT or CT imaging, and radiation therapy, including hadrontherapy. The end user in our case is not the medical physicist sending a treatment plan to be computed, but the researcher who provides a file of macros describing a simulation. The research community behind the GATE simulator needs to execute various GATE simulations easily and rapidly without worrying about the computing resources provided by the European Grid Infrastructure. This section will describe more into detail the GATE-Lab, its architecture and advanced features. To begin with, it will shortly present the underlying grid infrastructure.

<sup>5</sup> <http://www.ogf.org/documents/GFD.167.pdf>

<sup>6</sup> <http://eu-acgt.org/news/newsletters/summer-2009/single-article/archive/2009/july/article/hope-hospital-platform-for-e-health.html>

<sup>7</sup> <https://espace.cern.ch/partnersite/workspace/faust/Shared%20Documents/FaustinRoman.WP22.D1.pdf>

### 3.1 Grid environment

The European Grid Infrastructure (EGI<sup>8</sup>) is currently the largest production grid worldwide providing more than 100,000 CPUs and several Petabytes of storage. This distributed computing infrastructure was built by projects (DataGrid, EGEE-I, -II and -III) spanning from 2002 to 2010 and is now supported in collaboration with National Grid Initiatives (NGIs). EGI is used on a daily basis by thousands of scientists organized in over 200 Virtual Organizations. It uses the gLite middleware (Laure et al. (2006)), which provides high level services for scheduling and running computational jobs, as well as for data and grid infrastructure management.

A User Interface (UI) is the initial point of access to the grid from which the user can be authenticated and authorized to use the grid resources. From the UI the user can submit and cancel jobs, query their status and retrieve their output. These tasks are taken into account by a Workload Management System (WMS), which queues the user requests and dispatches them to the different computing centres available. The gateway to each computing centre is one or more Computing Element (CE) that will distribute the workload over the Worker Nodes (WN) i.e. the computing units available at this center. Files are stored on Storage Elements (SEs) and registered in the File Catalogs, which permit to locate files (or replicas) distributed on the grid. Data Management services are responsible for locating, replicating and accessing the data transparently.

### 3.2 Architecture

The GATE-Lab has been integrated in the porting and execution platform supporting grid applications in use at the Creatis laboratory. As illustrated in Fig. 1, the platform complies to a three-tier architecture composed of (i) the client which consists in the VBrower (Olabarriaga et al. (2006)) offering a GUI for grid data management (ii) a lab server managing task submission, monitoring and error handling and (iii) the grid itself, externally administrated and accessible through gLite.

The user interacts with the GATE-Lab client, which is a VBrower plugin designed specifically for GATE simulations and using the VBrower as a GUI for grid data management. The GATE-Lab performs automatic parameter checking, input files bundling and uploading, simulation submission and history management. The lab server hosts the DIANE pilot-job framework (Moscicki (2003a)) and the MOTEUR workflow engine (Glatard et al. (2008)). MOTEUR generates grid tasks from the GATE workflow description and submits them to a DIANE master using a generic task manager. According to the number of waiting tasks, the agent controller submits pilot jobs to the grid Workload Management System (WMS). Once running, DIANE pilots download and execute tasks on the grid worker nodes (WN), periodically uploading standard output and standard error to the master.

#### 3.2.1 The client

The GATE-Lab main tab allows for preparing a new simulation and launching it on the grid. The end user is asked for a name for the new simulation and for its main macro file. Starting from this main macro file, the client checks the simulation parameters, looks for all local inputs files, bundles them into an archive and uploads it to the grid. The end user is also asked for an estimation of the total CPU time of the simulation. Depending on this estimation, the simulation is split into a different number of tasks, so that the parallelization and therefore the

---

<sup>8</sup> <https://www.egi.eu>

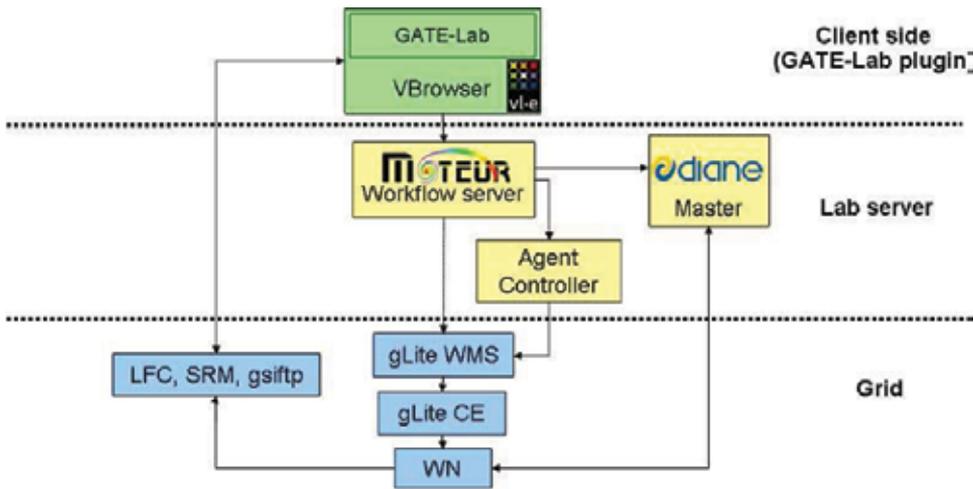


Fig. 1. Grid execution environment. On the client side, VBrower and the GATE-Lab plugin offer GUI for data management and launching GATE simulations on the grid. The lab server hosts the MOTEUR workflow engine and DIANE pilot-job framework. MOTEUR enacts application workflow and submits tasks to the DIANE master offering pilot-job execution. The agent controller submits pilot jobs to the grid.

speedup should be improved. At this point, a *one-click* GATE simulation is made possible for the user who does not have to be aware of grid internals as shown in Fig 2.

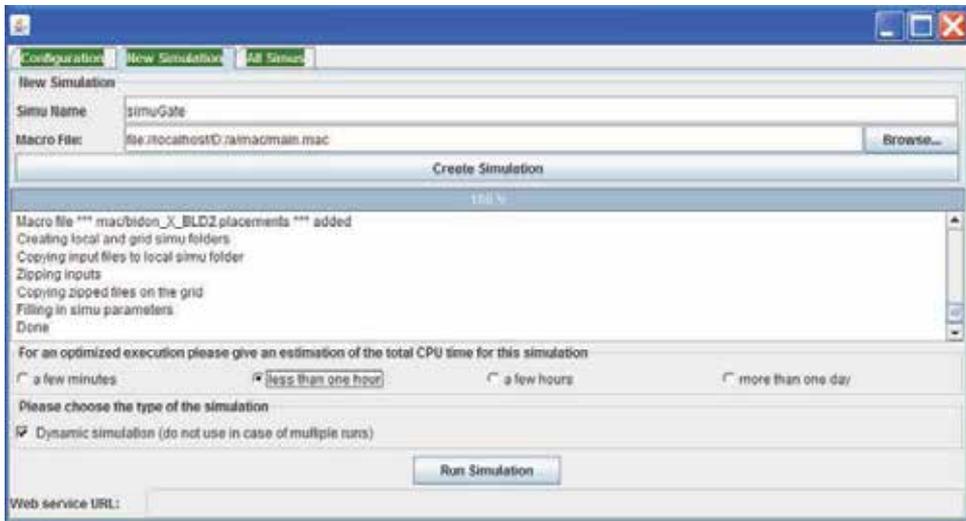


Fig. 2. GATE-Lab - create new simulation. Starting from the main macro file, the client checks the simulation parameters, looks for all local inputs files, bundles them into an archive and uploads it to the grid. Running GATE simulations on the grid is possible with a few clicks.

The number of parallel tasks to submit to the grid is automatically determined based on the end user estimation of the total CPU time needed for the complete simulation. The total CPU time depends on the number of events, but also on their type. Therefore, it is difficult to



Fig. 3. GATE-Lab - simulation monitoring. Simulation status is monitored in real time and outputs are accessible with an URL.

automatically forecast and the user who submits the simulation is the best person to estimate it. Nevertheless, he or she does not need to have the grid knowledge allowing to choose the right number of tasks for the application parallelization. The GATE-Lab integrates a mapping between the time estimation given by the end user and the number of tasks to submit. At the moment, the mapping corresponds to a minimum of 5 jobs for short simulations (a few minutes) and goes to a maximum of 500 jobs for simulations lasting more than a day.

The GATE-Lab implements two parallelization methods. The dynamic method gives the best results and is used by default. Nevertheless it cannot be used with all types of GATE simulations like for example simulations with multiple runs. In the Geant4 toolkit (Allison et al. (2006)) on which Gate is based, a "run" is the largest unit of simulation and consists of a sequence of "events" (particles history). Simulations dealing with time varying geometry perform successive runs. Therefore the end user has the possibility to use the standard static splitting method by simply deactivating the corresponding box (see Fig 2). More detail on these splitting methods will be given in section 3.3.1.

The GATE-Lab client has two more tabs: one for simulation history and the other for configuration purposes. The simulation history tab gives the list of links to all simulations launched by the user. The end user can thus easily retrieve his previous simulations, browse their web page or delete them. The configuration tab is important for allowing the user to choose certain parameters like for example the GATE release. Indeed, this was a strong requirement of the research community who needs to test new releases or switch to older ones for tests.

Once the GATE simulation is launched, its status is monitored and the end user can follow it in real time. The monitoring interface is presented in Fig 3. It is based on a PHP script available to the end user either within the VBrowser or from a simple web browser.

When the total number of events has been reached, a merging job is automatically launched. It merges all outputs generated in parallel and produces a final result easily accessible to the end-user through the VBrowser.

### 3.2.2 The Lab server

The lab server hosts the DIANE pilot-jobs framework and the MOTEUR workflow engine. MOTEUR generates grid tasks from the GATE workflow description and submits them to a DIANE master using a generic task manager. Pilot-job frameworks provide late-task binding to resources. Thus, tasks are no longer pushed to computing resources but generic pilots are submitted. Once running, pilots connect back to a central pool, fetching tasks when available and dying otherwise. Pilots are submitted on the grid with standard gLite command-line tools. This architecture is presented in Figure 4.

The GATE executable is deployed "on the fly" by the pilot-jobs. In order to do so, a GATE "release" is prepared (once and for all) for each new GATE version and stored on grid SEs. The release contains the GATE executable (compiled on a compatible architecture) and the necessary shared libraries. For each GATE task, pilots download the GATE release as well as all input files previously stored on grid SEs. To limit file transfers, a cache system has been implemented in pilots. Pilots cache the release and the inputs on the WN in case their next task requires the same files.

The agent controller monitors the number of pilots and decides of their submission according to the number of tasks created by MOTEUR. This mechanism, together with the indication on the total CPU time given by the end user, allows to modulate the number of resources asked/used depending on the size of the simulation.

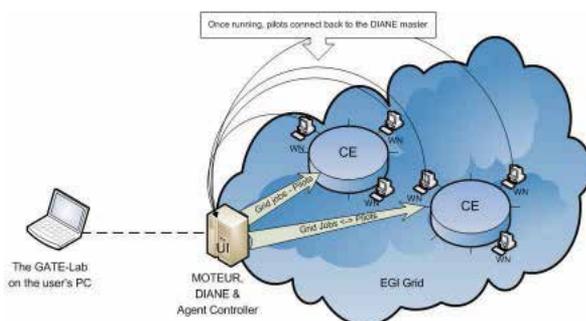


Fig. 4. DIANE pilot-jobs framework integrated into the Creatis grid execution environment.

Error handling is implemented by MOTEUR by resubmitting tasks up to a maximal number of times when they fail. Faulty pilots (i.e. pilots running a task that fails) are also removed and resubmitted if necessary (if no other 'free' pilots are available).

Authentication is based on X509 grid certificates. Short lived (maximum of 24h) grid proxies are generated with the VBrowser based on the user certificate. For GATE simulations longer than 24h, users have to renew their proxy. To ensure basic security, user credentials are delegated to the MOTEUR server, which starts a new DIANE master for every user.

### 3.3 Advanced features

The GATE-Lab is more than a friendly user interface for running GATE simulation on the grid. It integrates a set of advanced features which enhance its performance.

#### 3.3.1 Dynamic parallelization

As presented in Section 2.1, static parallelization methods tend to lead to poor scheduling mainly because of the heterogeneity of the grid resources. The GATE-Lab offers two parallelization options that have proved to be well adapted to the grid usage (Camarasu-Pop

et al. (2010)). Both methods use pilot jobs, which bring considerable advantages in terms of latency reduction and fault-tolerance on the grid (Germain Renaud et al. (2008)).

The first method is an intermediate solution based on a dynamic distribution (and reassignment) of statically partitioned tasks. In this case, a simulation is statically divided into a large number of tasks (of a convenient granularity) that are dynamically distributed to available resources. This method gives reasonably good results but underexploits the available resources, especially towards the end of the simulation.

Therefore a new dynamic task partitioning strategy was proposed and implemented to balance the number of particles that each resource has to simulate depending on its performance. The proposed dynamic task partitioning consists in a *do-while* algorithm with no initial splitting. Each computing resource contributes to the whole simulation until it is completed. As summarized in section 4 and detailed in Camarasu-Pop et al. (2010), the dynamic parallelization brings significant performance improvement, the makespan being on average twice smaller than with the previous method. Indeed a significant amount of time is saved on the completion of the last tasks. However this method cannot be used for all GATE simulations at the moment. GATE simulations with multiple runs for example need to respect additional criteria when parallelized. For these simulations the first method has to be used. Therefore, both methods are available in the GATE-Lab.

### 3.3.2 Stop and merge

GATE simulations can be very long and end-users may want to stop them beforehand if they consider that the number of already simulated particles is large enough and that the needed uncertainty level has been reached. Moreover, as previously mentioned, it is the last part of the simulation that is usually the slowest one. Therefore a "stop and merge" feature has been added to the GATE-Lab. It allows end users to stop the simulation at will and trigger the merging phase. This stop feature is possible due to an extension of the GATE code which is also used for the dynamic parallelisation. The GATE code was adjusted to handle stop signals during simulation. This add-on allows the application to pause regularly and check if the stop signal has been received. If this is the case, GATE saves its results and stops. Once all parallel results are saved and uploaded on the grid storage elements, the merging phase is launched to produce the final result.

This "stop and merge" feature raises a new challenge. Between the moment when the end-user sends the stop signal and the moment when the partial results are uploaded on the storage elements of the grid, job failures can happen. This will result in having less simulated particles than the user believed to have when he decided to stop the simulation. In a regular situation, failed jobs are resubmitted in order to ensure 100% complete results. In the situation of a stop and merge demand, the end-user is probably not willing to wait for job resubmission and wants the final result right away if the number of lost particles is not too important. Currently jobs that fail after the stop signal are not resubmitted. Nevertheless, an interaction with the end-user will have to be considered. The user will be presented the real status of the simulation and will have to make a second decision based on this new, more accurate information.

### 3.3.3 Incremental merge

Merging the partial results achieved in parallel is essential for the completion of the GATE simulation. This is a critical and delicate task because of its data intensive character. Indeed, partial results are stored on geographically distributed (and sometimes unreliable) grid

storage elements. The merging phase has the difficult task of downloading all partial results (currently up to 500 tarballs of a few tens of MB each) on one machine. Transfers are often long and may sometimes fail, thus endangering the success of the whole simulation as can be seen in the Results section.

Therefore, to improve the overall success rate and the performance of the merging phase, an incremental merge was implemented and is currently evaluated. In general, the chances to successfully retrieve a partial result are maximal shortly after it has been uploaded. The incremental merge does not wait for all results to be ready. It is launched along with the simulation as soon as a given number  $N$  (for example 5) of partial results have been produced. It merges these results into an intermediate result that will be in turn merged with the next partial or intermediate results. Thus, partial results are downloaded soon after they are produced, limiting the risk that they become unavailable. Moreover, a file that is temporarily unavailable has more chances to be retrieved by one of the successive instances of the incremental merge.

#### 4. Results

The system has been running since 2009 but has been more intensely used from April 1st to September 17th 2010 by two radiotherapy researchers, exploiting the resources supporting the biomed Virtual Organization of the EGEE grid. These gather 200+ CEs spread over 50 countries and are continuously shared by some 100 users. A detailed snapshot of the infrastructure usage is available on accounting portals<sup>9</sup>. A total of 197 simulations were submitted among which 89 were completely successful, 42 were “half-successful” (i.e. only some jobs completed) and 66 completely failed (no job completed). Tables 1 and 2 summarize performance results.

The elapsed time is the time perceived by the user, i.e., the duration between the launching of the simulation and the completion of the last task. The CPU time is a cumulative value on all the tasks of the simulation. It is an indication of the time that the simulation would have required on a single machine representative of the average grid node. The speed-up factor is a ratio between the CPU time and the elapsed time. The data/computing ratio is computed as the cumulative data transfer time (download + upload) divided by the cumulative CPU time. It gives an indication on the efficiency of the grid deployment. The error rate is computed as the ratio between the number of failed tasks (total tasks - successful tasks) and the total number of tasks.

Among the successful simulations (Table 1), a total of 8.4 CPU years were consumed and 27,798 grid tasks were submitted. Sixteen (16) simulations were slowed down by the grid execution. These were small test 6-job simulations, which suggests that a better handling of test simulations (e.g. by local execution) would be useful. The speed-up of the remaining simulations ranges from 3.12 to 146 with an overall value of 47. Although this value seems of little significance given the amount of available resources on the grid, it has to be noticed that this performance was obtained in production, on a shared platform. It is a sustainable level of performance that can be delivered to a group of users on a daily basis. As expected, there is some positive correlation between the speed-up and the number of submitted jobs, which suggests that the execution time choices made by the users through the GATE-Lab interface make sense.

---

<sup>9</sup> <http://www3.egee.cesga.es/>

	Elapsed	CPU	Speed -up	Data/ Computing	Total tasks	Error rate
min	397s	26s	0	0	6	0
max	3.9 days	174 days	146	16.65	815	0.61
average	0.73 days	34 days	-	-	312	-
all simulations	65 days	8.4 years	47	0.01	27798	0.09

Table 1. Performance figures of successful simulations (45% of all GATE-Lab simulations).

	Elapsed	CPU	Speed -up	Data/ Computing	Total tasks	Error rate
min	261s	26s	0	0	2	0
max	2 days	200 days	369	9.07	977	1
average	0.69 days	54 days	-	-	373	-
all simulations	31 days	6.5 years	77	0.01	16417	0.33

Table 2. Performance figures of half-successful simulations (some completed jobs, 21% of all GATE-Lab simulations).

The overall data/computing ratio is 1%, which is a very good value on wide scale platforms such as the EGI. Again it shows that the number of jobs was properly chosen. High data/computing values are observed for 6-job simulations though and few outliers are also seen for 502- or 256-job simulations. These probably come from temporary network issues or overloads on the storage system.

The average job error ratio is 9%, which is a good value on the EGI. Most of the errors come from data transfer issues. Other sources of errors include pilot-master communication problems (lost pilots) and application issues coming from the configuration of the worker nodes.

Half-successful simulations (Table 2) did not complete due to some repeated failures in simulation or merge tasks. Some of these simulations could still be exploited thanks to manual recovery, in particular in case of merge failures. As expected, the overall error ratio (33%) is far greater than for successful simulations. The speed-up is also higher than for successful simulations due to the lack of merge task.

Failed simulations represented 26 days of elapsed time and a total of 2074 grid tasks. These failures are due either to critical problems of the infrastructure (GATE-Lab or grid) or to user mistakes. In the future, mechanisms must be envisaged to detect these issues earlier in the simulation.

Figure 6 shows an example of task flow obtained with static respectively dynamic partitioning of a GATE simulation. The dynamic partitioning obviously leads to a better scheduling than the static. Thanks to the lack of resubmission and better resource exploitation, no slowdown is observed towards the end of the simulation for the dynamic approach. Figure 5 shows the simulation completion along time for the dynamic (red curves) and static (black curves) parallelization modes. During the first phase of the simulation (up to 400,000 particles), the dynamic parallelization shows a better throughput (number of simulated particles per seconds) but the overall time gain remains modest. More importantly, the dynamic parallelization dramatically improves the performance of the simulation during its last phase (from 400,000 simulated particles to 450,000). This highlights the benefit obtained using our load-balancing method.

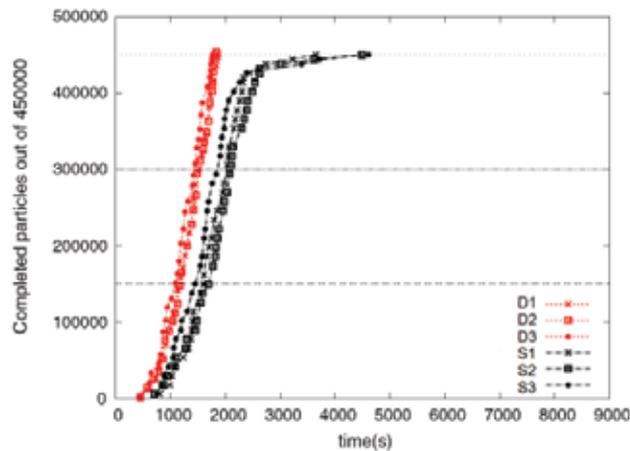


Fig. 5. GATE simulation completion using the static VS the dynamic parallelization approach. When using the static method the last tasks are considerably slowing down the simulation (figure extracted from Camarasu-Pop et al. (2010)).

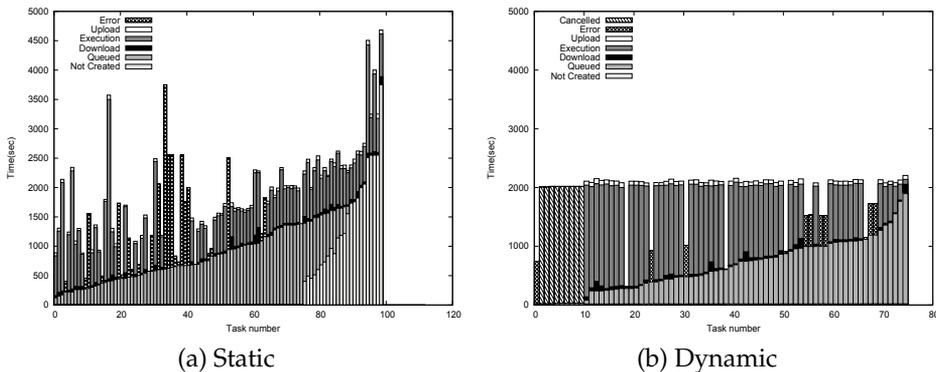


Fig. 6. Example of task flow obtained with (a) Static & (b) Dynamic partitioning of a GATE simulation on EGI with 75 submitted pilots. Available resources are all exploited until the end of the simulation. Errors are compensated without any pilot resubmissions. The dynamic scheduling obviously outperforms the one obtained with static partitioning (figure extracted from Camarasu-Pop et al. (2010)).

## 5. Discussion / Lessons learned

The GATE-Lab has been created to serve the GATE (radiotherapy) research community. The main challenge of the project was to make a sustainable service from a research code running on a distributed environment. The difficulty of this task resided not only in the grid complexity and its rendering transparent to the research community, but also in integrating in the system an evolving research code.

Concerning the execution of GATE on grid nodes, we learned that it was important to have "application administrators" among GATE researchers. They are at the interface between grid and GATE experts ensuring thus a good communication between the two communities. As an opensource software developed by the international OpenGATE collaboration, GATE is always evolving and new releases are available periodically. In order to run on the grid,

these new releases must be compiled and packaged properly for the grid environment. GATE application administrators are in charge of preparing these releases with the help of the tools provided by grid experts.

Regarding the grid environment, we grew aware of the importance of using pilot jobs. Pilot jobs bring dramatic improvement w.r.t. default gLite, both in terms of reliability and performance. gLite standard submission never manages to complete 100% of the simulation, the success rate being of 75% on average. Workflows are also very useful for application porting and reusability. A middleware-independent workflow description of the application simplifies migration to other execution frameworks. In our case, we can easily extend GATE submission to platforms where pilots are not an option or migrate to other pilot frameworks. In spite of all the efforts put into the optimization and automation of the GATE-Lab, user support is needed regularly. Most of the time support is needed for the merging phase and other unexpected error handling like for example temporary failures of grid services.

Last but not least, the merging phase represents a real challenge because of its data intensive character and its performance needs to be improved. From our experience we know that this merging activity and related problems are not specific to GATE; they are common to many applications parallelized on the grid.

## 6. Conclusion

This chapter presented a solution for reliable, user-friendly and fast execution of GATE (an open-source software for nuclear medicine and radiation therapy simulations) on the grid by using the GATE-Lab. The GATE-Lab targets researchers wishing to easily execute GATE with different sets of parameters which do not necessarily correspond to the treatment plans used in clinic.

The GATE-Lab plugin is build upon the porting and execution platform supporting grid applications in use at the Creatis laboratory. The plugin has been specifically designed to answer researchers' requirements. The current release is the result of a mature project started two years ago and which evolved according to the researchers' feedback. Advanced features like the stop on merge and the incremental merge have been added at the researchers' request. During the last 5 months the system provided an average speed-up of almost 50. Although not outstanding, this is a good level of performance since it was delivered on a shared infrastructure, when several other users from our lab and others were also running heavy computations. Most of the simulations were successful but it was noticed that the support for test and prototyping simulations was not very efficient and that user mistakes leading to numerous task failures should be better detected.

The dynamic parallelization is a powerful feature that has been proposed and implemented for GATE. It is a general Monte-Carlo load-balancing method that could greatly speed-up other kinds of Monte-Carlo simulations.

The GATE-Lab is already intensively used by the (non-clinical) radiation therapy researchers at Creatis. As future work we plan to make it available for other research teams worldwide. Within the hGate project we also plan to extend the GATE-Lab to other computing platforms (e.g. local clusters).

## 7. Acknowledgments

This work is co-funded by the French national research agency (ANR), hGATE project under contract number ANR-09-COSI-004-01. It also falls into the scope of the scientific topics of the

French National Grid Institute (IdG). The authors would like to thank the site administrators of the European Grid Initiative and the GGUS support for their work and Fabrice Bellet for his help with application compiling and customization for the grid. We also would like to thank Piter T. de Boer and Spiros Koulouzis (NIKHEF) for the support with the VBrowser, Jakub T. Mosciki (CERN) for the help with DIANE and the Modalis team (CNRS, I3S) for the MOTEUR developments.

## 8. References

- Ahn, S., Namgyu, K., Seehoon, L., Soonwook, H., Dukyun, N., Koblitz, B., Breton, V. & Sangyong, H. (2008). Improvement of Task Retrieval Performance Using AMGA in a Large-Scale Virtual Screening, *NCM'08*, pp. 456–463.
- Allison, J., Amako, K., Apostolakis, J., Araujo, H., Dubois, P., Asai, M., Barrand, G., Capra, R., Chauvie, S., Chytracsek, R., Cirrone, G., Cooperman, G., Cosmo, G., Cuttone, G., Daquino, G., Donszelmann, M., Dressel, M., Folger, G., Foppiano, F., Generowicz, J., Grichine, V., Guatelli, S., Gumplinger, P., Heikkinen, A., Hrivnacova, I., Howard, A., Incerti, S., Ivanchenko, V., Johnson, T., Jones, F., Koi, T., Kokoulin, R., Kossov, M., Kurashige, H., Lara, V., Larsson, S., Lei, F., Link, O., Longo, F., Maire, M., Mantero, A., Mascialino, B., McLaren, I., Lorenzo, P., Minamimoto, K., Murakami, K., Nieminen, P., Pandola, L., Parlati, S., Peralta, L., Perl, J., Pfeiffer, A., Pia, M., Ribon, A., Rodrigues, P., Russo, G., Sadilov, S., Santin, G., Sasaki, T., Smith, D., Starkov, N., Tanaka, S., Tcherniaev, E., Tome, B., Trindade, A., Truscott, P., Urban, L., Verderi, M., Walkden, A., Wellisch, J., Williams, D., Wright, D. & Yoshida, H. (2006). Geant4 Developments and Applications, *IEEE Transactions on Nuclear Science* 53: 270–278.
- Bagnasco, S., Betev, L., Buncic, P., Carminati, F., Cirstoiu, C., Grigoras, C., Hayrapetyan, A., Harutyunyan, A., Peters, A. J. & Saiz, P. (2008). Alien: Alice environment on the grid, *Journal of Physics: Conference Series* 119(6).
- Camarasu-Pop, S., Benoit-Cattin, H., Guigues, L., Clarysse, P., Bernard, O. & Friboulet, D. (2008). Towards a virtual radiological platform based on a grid infrastructure, *Medical imaging on grids: achievements and perspectives (MICCAI Grid Workshop)*, New York, USA, pp. 85–95.
- Camarasu-Pop, S., Glatard, T., Mościcki, J., Benoit-Cattin, H. & Sarrut, D. (2010). Dynamic partitioning of GATE Monte-Carlo simulations on EGEE, *Journal of Grid Computing*.
- Cirne, W., Brasileiro, F., Paranhos, D., Goes, L. & Voorsluys, W. (2007). On the efficacy, efficiency and emergent behavior of task replication in large distributed systems, *Parallel Computing* 33: 213–234.
- Compostella, G., Lucchesi, D., Griso, S. P. & Sfiligoi, I. (2007). CDF Monte Carlo Production on LCG Grid via LcgCAF Portal, *E-SCIENCE '07: Proceedings of the Third IEEE International Conference on e-Science and Grid Computing*, IEEE Computer Society, Washington, DC, USA, pp. 11–16.
- Deelman, E., Singh, G., Su, M.-H., Blythe, J., Gil, Y., Kesselman, C., Mehta, G., Vahi, K., Berriman, G. B., Good, J., Laity, A., Jacob, J. C. & Katz, D. S. (2005). Pegasus: a Framework for Mapping Complex Scientific Workflows onto Distributed Systems, *Scientific Programming Journal* 13(3): 219–237.
- Engh, D., Smallen, S., Gieraltowski, J., Fang, L., Gardner, R., Gannon, D. & Bramley, R. (2003). GRAPPA: Grid access portal for physics applications, *CoRR* cs.DC/0306133.
- Galyuk, Y. P., Memnonov, V., Zhuravleva, S. E. & Zolotarev, V. I. (2002). Grid technology with dynamic load balancing for Monte Carlo simulations, *PARA '02: Proceedings of the 6th*

- International Conference on Applied Parallel Computing Advanced Scientific Computing*, Springer-Verlag, London, UK, pp. 515–520.
- Germain Renaud, C., Loomis, C., Moscicki, J. & Texier, R. (2008). Scheduling for Responsive Grids, *Journal of Grid Computing* 6: 15–27.
- Glatard, T., Montagnat, J., Lingrand, D. & Pennec, X. (2008). Flexible and efficient workflow deployment of data-intensive applications on grids with MOTEUR, *International Journal of High Performance Computing Applications (IJHPCA)* 22(3): 347–360.
- Jacq, N., Salzemann, J., Jacq, F., Legré, Y., Medernach, E., Montagnat, J., Maass, A., Reichstadt, M., Schwichtenberg, H., Sridhar, M., Kasam, V., Zimmermann, M., Hofmann, M. & Breton, V. (2008). Grid enabled virtual screening against malaria, *Journal of Grid Computing* 6: 29–43.
- Jan, S., Santin, G., Strul, D., Staelens, S., AssiÉ, K., Autret, D., Avner, S., Barbier, R., Bardières, M., Bloomfield, P. M., Brasse, D., Breton, V., Bruyndonckx, P., Buvat, I., Chatziioannou, A. F., Choi, Y., Chung, Y. H., Comtat, C., Donnarieix, D., Ferrer, L., Glick, S. J., Groiselle, C. J., Guez, D., Honore, P. F., Kerhoas-Cavata, S., Kirov, A. S., Kohli, V., Koole, M., Krieguer, M., van der Laan, D. J., Lamare, F., Largeron, G., Lartizien, C., Lazaro, D., Maas, M. C., Maigne, L., Mayet, F., Melot, F., Merheb, C., Pennacchio, E., Perez, J., Pietrzyk, U., Rannou, F. R., Rey, M., Schaart, D. R., Schmidtlein, C. R., Simon, L., Song, T. Y., Vieira, J. M., Visvikis, D., de Walle, R. V., Wieërs, E. & Morel, C. (2004). GATE: a simulation toolkit for PET and SPECT., *Phys Med Biol* 49(19): 4543–4561.
- José Carlos Mouriño Gallego, Andrés Gómez C. F. S. F. J. G. C. D. A. R. S. J. P. G. F. G. R. D. G. C. M. P. C. (2007). *1st Iberian Grid Infrastructure Conference Proceedings (IBERGRID)*.
- Kacsuk, P., Farkas, Z. & Fedak, G. (2008). Towards making BOINC and EGEE interoperable, *4th eScience Conference*, Indianapolis, pp. 478–484.
- Kacsuk, P. & Sipos, G. (2005). Multi-Grid, Multi-User Workflows in the P-GRADE Grid Portal, *Journal of Grid Computing (JGC)* 3(3-4): 221 – 238.
- Laure, E., Fisher, S., Frohner, A., Grandi, C., Kunszt, P., Krenek, A., Mulmo, O., Pacini, F., Prelz, F., White, J., Barroso, M., Buncic, P., Byrom, R., Cornwall, L., Craig, M., Meglio, A. D., Djaoui, A., Giacomini, F., Hahkala, J., Hemmer, F., Hicks, S., Edlund, A., Maraschini, A., Middleton, R., Sgaravatto, M., Steenbakkens, M., Walk, J. & Wilson, A. (2006). Programming the Grid with gLite, *Computational Methods in Science and Technology* 12(1): 33–45.
- Maeno, T. (2008). Panda: distributed production and distributed analysis system for atlas, *Journal of Physics: Conference Series* 119(6): 062036 (4pp).
- Maheshwari, K., Missier, P., Goble, C. & Montagnat, J. (2009). Medical Image Processing Workflow Support on the EGEE Grid with Taverna, *Intl Symposium on Computer Based Medical Systems(CBMS'09)*, IEEE.
- Maigne, L., Hill, D., Calvat, P., Breton, V., Lazaro, D., Reuillon, R., Legré, Y. & Donnarieix, D. (2004). Parallelization of Monte-Carlo simulations and submission to a grid environment, *Parallel Processing Letters HealthGRID 2004*, Vol. 14, Clermont-Ferrand France, pp. 177–196.
- M.G. Bugeiro, J.C. Mouriño, A. G. C. V. E. H. I. L. y. D. R. (2009). Integration of slas with gridway in beineimrt project, *3rd Iberian Grid Infrastructure Conference*.
- Montagnat, J., Breton, V. & Magnin, I. (2005). Partitionning medical image databases for content-based queries on a grid, *Methods of Information in Medicine (MIM)* 44(2): 154–160.

- Moscicki, J. T. (2003a). Diane - distributed analysis environment for grid-enabled simulation and analysis of physics data, *Nuclear Science Symposium Conference Record, 2003 IEEE*, Vol. 3, pp. 1617–1620 Vol.3.
- Moscicki, J. T. (2003b). Distributed analysis environment for HEP and interdisciplinary applications, *Nuclear Instruments and Methods in Physics Research A* 502: 426–429.
- Oinn, T., Addis, M., Ferris, J., Marvin, D., Senger, M., Greenwood, M., Carver, T., Glover, K., Pocock, M. R., Wipat, A. & Li, P. (2004). Taverna: A tool for the composition and enactment of bioinformatics workflows, *Bioinformatics journal* 17(20): 3045–3054.
- Olabarriaga, S., de Boer, P. T., Maheshwari, K., Belloum, A., Snel, J., Nederveen, A. & Bouwhuis, M. (eds) (2006). *Virtual Lab for fMRI: Bridging the Usability Gap*, IEEE, Amsterdam.
- Olabarriaga, S., Glatard, T. & de Boer, P. T. (2010). A virtual laboratory for medical image analysis, *IEEE Transactions on Information Technology In Biomedicine (TITB)*.
- Pena, J., González-Castaño, D. M., Gomez, F., Gago-Arias, A., González-Castaño, F. J., Rodríguez-Silva, D. A., González, D., Pombar, M., Sánchez, M., Portas, B. C., Gómez, A. & Mouriño, C. (2009). E-IMRT: a web platform for the verification and optimization of radiation treatment plans., in R. Magjarevic, J. H. Nagel, O. Dössel & W. C. Schlegel (eds), *World Congress on Medical Physics and Biomedical Engineering, September 7 - 12, 2009, Munich, Germany*, Vol. 25/1 of *IFMBE Proceedings*, Springer Berlin Heidelberg, pp. 511–514.
- Procassini, R., O'Brien, M. & Taylor, J. (2005). Load Balancing of Parallel Monte Carlo Transport Calculations, *Mathematics and Computation, Supercomputing, Reactor Physics and Nuclear and Biological Applications*, Palais des Papes, Avignon, Fra.
- Reuillon, R., Hill, D., Gouinaud, C., El Bitar, Z., Breton, V. & Buvat, I. (2008). Monte Carlo Simulation With The GATE Software Using Grid Computing, *Proceedings of NOTERE 2008 8ème Conférence Internationale sur les NOuvelles TEchnologies de la REpartition, NOTERE 2008*, Lyon France.
- Sfiligoi, I. (2008). glideinWMS - A generic pilot-based workload management system, *Journal of Physics: Conference Series* 119(6): 062044 (9pp).
- Tan, W.-J., Ching, C. T. M., Camarasu-Pop, S., Calvat, P. & Glatard, T. (2010). Two experiments with application-level quality of service on the egee grid, *GMAC '10: Proceeding of the 2nd workshop on Grids meets autonomic computing*, ACM, New York, NY, USA, pp. 11–20.
- Tsaregorodtsev, A., Bargiotti, M., Brook, N., Ramo, A. C., Castellani, G., Charpentier, P., Cioffi, C., Closier, J., Diaz, R. G., Kuznetsov, G., Li, Y. Y., Nandakumar, R., Paterson, S., Santinelli, R., Smith, A. C., Miguelez, M. S. & Jimenez, S. G. (2008). Dirac: a community grid solution, *Journal of Physics: Conference Series* 119(6): 062048 (12pp).

# Monte Carlo Simulation for Ion Implantation Profiles, Amorphous Layer Thickness Formed by the Ion Implantation, and Database Based on Pearson Function

Kunihiro Suzuki  
*Fujitsu Laboratories Ltd.*  
*Japan*

## 1. Introduction

Ion implantation is a standard technology to dope substrates in Si VLSI processes. The ion implantation profiles in Si substrates are generated based on a vast secondary ion mass spectrometry (SIMS) database of ion implantation profiles in commercial simulators such as Sentaurus Process. However, we cannot predict profiles accurately when there are no experimental data or only poor data. Recently, various ions such as C, N, and F have been used to suppress transient enhanced diffusion in the subsequent annealing processes, as shown by Hu (2000) and Mirabera (2005). Furthermore, various substrates have been investigated, such as SiGe for state of the art Si LSI [Kim (2006), Weber (2007)], Ge for post Si devices [Chui (2002), Shang (2003)], and SiC for high-temperature, high-voltage, and high-power applicants [Schoerner (1999)]. Ion implantation is also a standard technology to dope these substrates. However, accumulation of the corresponding experimental database is time and cost consuming. Therefore, theoretical evaluation of these profiles is invoked.

Monte Carlo (MC) simulation is widely used for predicting ion implantation profiles and has long been developed to be available for any combination of incident ion and substrate atoms [Ziegler (2008) SRIM, Tian (2003), Suzuki (2009)]. We can evaluate the accuracy of the MC by comparing an ion implantation database such as FabMeister-IM [Suzuki (2010a)]. We show that the MC results sometimes deviate from the experimental data with its original form. We modified the electron stopping power model, calibrated its parameters, and reproduced most of the database with the energy range between 0.5 and 2000 keV.

MC simulation takes long time to calculate the profiles, and the profiles are scattered in the low concentration region. Therefore, MC results are sometimes converted to an analytical Pearson function presented by Hofker (1975) and Ashworth (1990) utilizing its moment parameters, with which we can expect smooth profiles over the entire region. Furthermore, we can use the moment parameters of MC as a database and can instantaneously obtain profiles using the Pearson function for various ion implantation conditions by interpolating the moment parameter values. However, we show that direct use of the moment parameters evaluated from MC data sometimes induces inaccurate Pearson function, as shown in

Suzuki (2010b). We show the way for obtaining moment parameters to reproduce MC results using Pearson function.

Amorphous layer is formed by the ion implantation. This amorphous layer is utilized to obtain shallow junctions by suppressing channelling effect. B ions implanted into substrates with and without an amorphous layer formed by Ge ion implantation are shown in Suzuki (2010a). If an amorphous layer is formed and ion implantation is used to add impurities into this amorphous layer, a significant shallow junction can be obtained with no channelling tail.

Furthermore, the B activation phenomenon becomes to be drastically changed where the amorphous layer is formed or not. The sheet resistance with annealing time at 600°C in substrates with an amorphous layer formed by Ge ion implantation is significantly reduced compared to that without forming an amorphous layer [Suzuki (2004)]. These results show that by forming an amorphous layer and allowing it to recrystallize at a low temperature, it is possible to obtain low resistivity over a wide time range. It has been confirmed that this phenomenon is not intrinsic to Ge and can also be observed in amorphous layers formed by other impurities where recrystallization is allowed to take place at temperatures that produce little redistribution of impurities [Solmi (1990), Jin (2002), Suzuki (2003), Pawlak (2005)]. Therefore, it is also very important to predict the amorphous layer thickness depending on various ion implantation conditions.

MC simulation also has information on energy transferred to the substrate atoms, that is, induced damage, and the damage has been related to the amorphous layer thickness [Hobler (1988), Cerva(1992)]. Prussin (1984) also analyzed the formation of the continuous amorphous layer based on Brice's energy deposition model [Brice (1975)]. We also show that the amorphous layer thickness can also be predicted by MC using a critical vacancy concentration.

## 2. Experiments

We deposited around 1  $\mu\text{m}$  of Si by low pressure chemical vapour deposition at 550°C on Si substrates or formed an amorphous layer by Ge ion implantation. We verified that the profiles near the peak and surface regions in these amorphous layers are almost the same as the profiles in crystal Si (cSi) substrates. Therefore, we also used the profiles in cSi, neglecting the channelling tail region to evaluate the accuracy of MC results.

We evaluated ion implanted impurity concentration profiles using SIMS. In the SIMS measurement, the primary ions were raster scanned over a wide area, and secondary ions were collected from the central small area using electronic gating to avoid edge effects. The depth calibration of the measured profile was done using a Dektak 2A surface profilometer, and the concentration scale was adjusted to the as-implanted dose. The standard SIMS measurement conditions are shown in ref. [Suzuki (1998)]. The accurate SIMS measurement for ultra shallow profiles has been developed by [Kataoka (2007), Tada (2008)], which enables us to compare SIMS and MC in a low ion implantation energy region of around 1 keV.

We also evaluated the amorphous layer thickness by transmission electron microscopy (TEM). TEM measurements were performed using a JEOL 2500 transmission electron microscope operating at 200 keV. Cross-sectional view specimens were obtained by ion milling [Suzuki (2006)].

### 3. Brief review of the physics of ion implantation

We will briefly explain the physics of ion implantation. A detailed description can be found in Ziegler (2008).

In a nuclear interaction, a binary interaction is assumed. The energy transferred from the incident atom to the target atom  $T_{2f}$  is given by

$$T_{2f} = \frac{4M_1M_2}{(M_1 + M_2)^2} \sin^2\left(\frac{\Phi}{2}\right) T_{1i}, \quad (1)$$

where  $T_{1i}$  is the incident atom energy, and  $M_1$  and  $M_2$  are the incident and target atom mass numbers, respectively.  $\Phi$  is the scattering angle calculated using Ziegler-Litmark-Biersack universal potential model [Ziegler (2008)].

On the other hand, Lindhard proposed an electron stopping power of

$$S_e = r_e 1.21 \times 10^{-16} Z_1^{1/6} \frac{Z_1 Z_2}{(Z_1^{2/3} + Z_2^{2/3})^{3/2}} \frac{1}{\sqrt{M_1 [g]}} \sqrt{E [eV]} \quad [eV \cdot cm^2], \quad (2)$$

where  $Z_1$  and  $Z_2$  are the incident and target atomic numbers, respectively [Lindhard (1961)].  $r_e$  is a fitting parameter. Lindhard's  $S_e$  model of Eq. 2 assumes the interaction between the electron cloud of ions and substrate atoms. However, the electron cloud of incident ions is expected to be stripped at high-energy regions. Therefore, Lindhard's model becomes invalid at high-energy regions.

Bethe derived an electron stopping power model, which is valid at high-energy regions, where the electron cloud is completely stripped as [Bethe (1930)]

$$S_e(E) = \left(\frac{q^2}{4\pi\epsilon_0}\right)^2 \frac{2\pi M_1 Z_2 Z_1^2}{m_e E} \ln\left(\frac{4m_e E}{I M_1}\right), \quad (3)$$

where  $q$  is the electron charge,  $\epsilon_0$  is the permittivity in vacuum,  $m_e$  is the electron mass, and  $I$  is the average electron excitation energy, and it is given in an empirical form as [Dalton (1968)]

$$I = \begin{cases} 11.2 + 11.7Z_2 & \text{for } Z_2 \leq 13 \\ 52.8 + 8.71Z_2 & \text{for } Z_2 > 13 \end{cases} \quad (4)$$

We propose to combine Lindhard's model with Bethe's model as follows.

First, Bethe's model is invalid in low-energy regions, and we modify it not to influence Lindhard's model in low-energy regions. The energy where Bethe's model has a maximum value can be evaluated from  $\frac{\partial S_e}{\partial E} = 0$ , and we obtain

$$E = eE_r, \quad (5)$$

where  $e$  is the base of natural logarithm, and

$$E_r = \frac{M_1}{4m_e} I. \quad (6)$$

We modify Bethe's model  $S_{e\_mB}(E)$  as

$$S_{e\_mB}(E) = \begin{cases} \left( \frac{q^2}{4\pi\epsilon_0} \right)^2 \frac{2\pi M_1 Z_2 Z_1^2}{m_e e E_r} & \text{for } E \leq eE_r, \\ \left( \frac{q^2}{4\pi\epsilon_0} \right)^2 \frac{2\pi M_1 Z_2 Z_1^2}{m_e E} \ln \left( \frac{E}{E_r} \right) & \text{for } E > eE_r, \end{cases} \quad (7)$$

Biersack et al. proposed a model similar to Eq. 7 with a mathematical trick [Biersack (1980)]. Both modified Bethe's models become the original Bethe's model at high-energy regions and much larger than Lindhard's model at low-energy regions. We propose to combine Lindhard's model and the modified Bethe's model of Eq. 7 as [Suzuki (2009)]

$$\frac{1}{S_e} = \left[ \left( \frac{1}{r_e S_{eL}} \right)^\theta + \left( \frac{1}{r_{eh} S_{e\_mB}} \right)^\theta \right]^{1/\theta} \quad (8)$$

We also introduce a fitting parameter  $r_{eh}$  for generality although we use  $r_{eh}$  of 1 in this analysis. (We need experimental data in the energy region much larger than  $eE_r$  to calibrate  $r_{eh}$ , which we have none here). This  $S_e$  becomes Lindhard's model at low-energy regions, and Bethe's model at high-energy regions. When  $\theta$  is one, it is the same form of the Biersack's model [Biersack (1980)].  $\theta$  empirically expresses the transition from Lindhard's model to Bethe's model.

Figure 1 shows the dependence of  $S_e$  of B in Si on  $\theta$ . The interaction between both models becomes significant with decreasing  $\theta$ . We can evaluate  $S_e$  of B by varying the value of  $\theta$  in the MeV energy region. It is also clear that we cannot evaluate robustly  $r_{eh}$  with the experimental data less than 10000 keV.

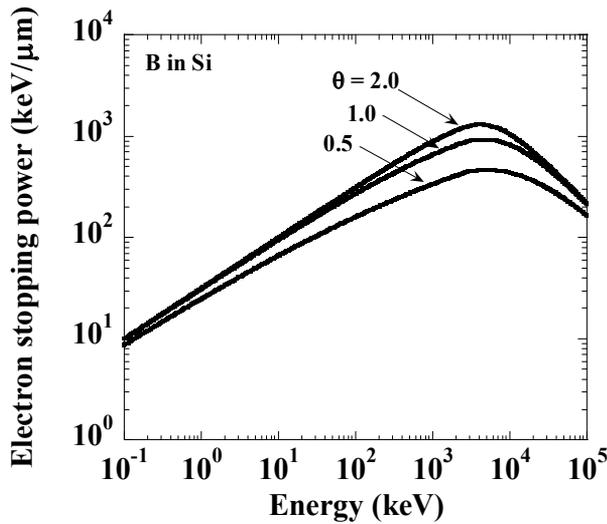


Fig. 1. Dependence of  $S_e$  of B in Si substrate on  $\theta$ .

We can evaluate the energy transferred to the substrate atom of  $T_{2f}$  given by Eq. 1. The recoiled atoms are generated if the  $T_{2f}$  is larger than the critical displacement energy  $E_d$ . We then trace trajectories of the recoiled atom with the energy of  $T_{2f} - E_d$  and count up the vacancies generated by the recoiled atom.

Modified Kinchin-Peace model is commonly applied to the primary recoiled atoms instead of tracing the recoiled substrate atom to save computation time [Ziegler (2008)]. The number of vacancy is evaluated using an analytical formula as a function of  $T_{2f}$ , and it is recorded at the location of the primary recoiled position. In this treatment, the damage (vacancy) can be expressed by

$$v = \begin{cases} 0.4 \frac{T_{2f}}{E_d} & \text{for } T_{2f} \geq 2.5E_d \\ \frac{T_{2f}}{E_d} & \text{for } E_d \leq T_{2f} < 2.5E_d \\ 0 & \text{for } T_{2f} < E_d \end{cases} \quad (9)$$

The transferred energy is not perfectly consumed by nuclear interaction, but some are consumed through the electron stopping power during many collisions. To ensure the assumption, the term of 0.4 is added in the first term of Eq. 9.  $E_d$  is an important parameter that control the radiation damage, but the experimental methods show widely varying results for  $E_d$  of Si substrate in the range of 10-30 eV, and theoretical evaluation shows that it depends on the direction of recoiled atom in the lattice and also in the range between 12 and 36 eV and average of around 24 eV [Holmstrom (2008)]. We used  $E_d = 25$  eV for both Si and Ge substrates in this analysis. This value influences the vacancy concentration, but the similar results can be expected using different  $E_d$  values.

We implemented the above physics in our own Monte Carlo simulator [Suzuki 2009]. We can switch two modes for evaluating the vacancy concentration: one is not tracing the recoiled substrate atom and use Eq. 10, and the other one is tracing the recoiled substrate atom trajectories. The former is called non-tracing mode, and the latter tracing mode is denoted by T.

#### 4. Comparison of MC with SIMS data

Figure 2 shows the comparison of SIMS and MC data with  $r_e$  of 1 for B, P, and As implantation. We obtained good agreement between MC and SIMS data for As profiles, and close agreement for P profiles, but significant deviation for B profiles. This may mean that we cannot expect predictive results from MC simulation since we cannot have clear physical reason whether our calculation for new ion in a certain substrate is like As or B profiles.

Figures 3 compares SIMS and MC B data with various  $r_e$ . The profile becomes shallower with increasing  $r_e$ . We can obtain close agreement of peak position as well as the overall shape of the profile with  $r_e$  of around 1.5.

Figure 4 compares various energy-dependent B SIMS profile data with the MC simulation with optimized  $r_e$  of 1.55. It is noteworthy that we can fit the data with a single  $r_e$  over a wide energy range.

We obtained a similar agreement for As SIMS data for various energies as shown in Fig. 5 using  $r_e$  of 1.0 as the same value in Fig. 2. The default value of  $r_e$  of 1 is also valid for various energies for As.

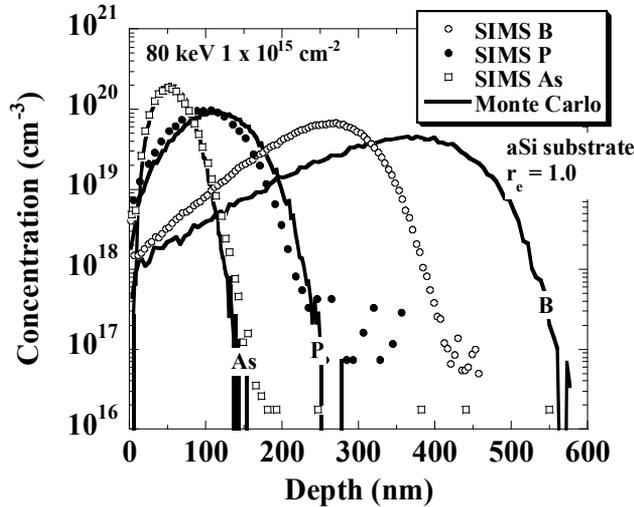


Fig. 2. Comparison of MC with SIMS data with  $r_e = 1$ .

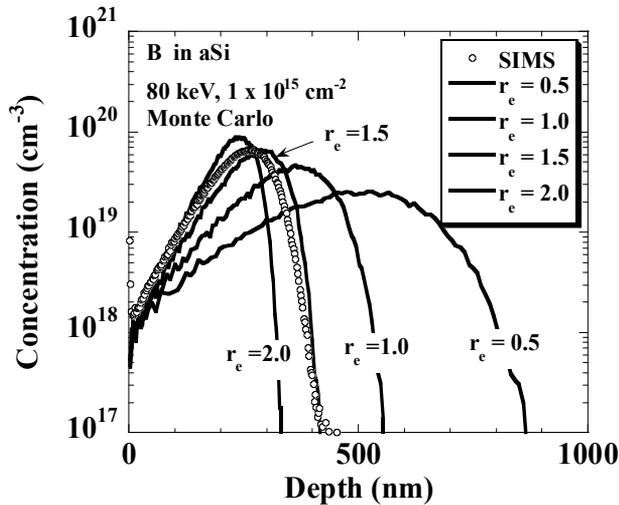


Fig. 3. Comparison of MC with B SIMS data at different values of  $r_e$ .

We obtained the similar results for the other ions of In, Sb, Ga, Ge, Si, N, F, and C in Si substrate, and B, P, and As in Ge substrate, and further in Mo, HfO<sub>2</sub>, and photo-resist substrates by tuning corresponding  $r_e$  that is valid for various energies, which is shown in [Suzuki (2009)]. Table 1 summarizes optimized  $r_e$ . The values of  $r_e$  are not far from 1 and it is 1 in many cases. Therefore, we can predict the profiles in an amorphous layer using a MC simulation with a default  $r_e$  deduced from the table, and we can further improve the accuracy if we tune  $r_e$  with some experimental data.

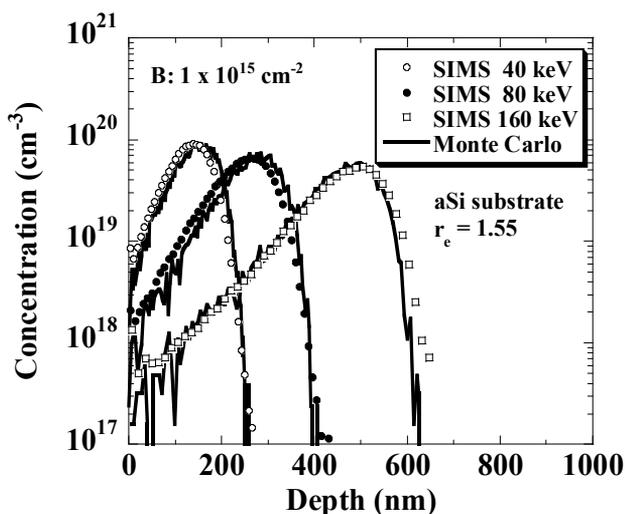


Fig. 4. Comparison of SIMS B profiles with MC with optimized  $r_e = 1.55$ .

		$Z_2$						
		01:H	06:C	08:O	14:Si	32:Ge	42:Mo	72:Hf
$Z_1$	05:B	1.0	1.7	0.8	1.55	1.0	1.0	3.0
	06:C	---	---	---	1.5	---	---	---
	07:N	---	---	---	1.4	---	---	---
	09:F	---	---	---	1.0	---	---	---
	14:Si	---	---	---	1.25	1.0	---	---
	15:P	1.0	1.0	1.0	1.2	1.0	1.0	1.0
	32:Ge	---	---	---	1.0	---	---	---
	33:As	---	---	0.5	1.0	1.0	---	1.0
	49:In	---	---	---	1.0	---	---	---
	51:Sb	---	---	---	1.0	---	---	---

Table 1.  $r_e$  for various incident ion and substrate atoms.

Low-energy ion implantation of around 1 keV is frequently used to realize shallow junctions. There is no critical point at this energy region from the standpoint of physics. However, SIMS reaches its resolution limit in this energy region. Therefore, the accuracy of the MC simulation in low-energy region has not been robustly evaluated. Recently, fundamental SIMS measurement mechanisms have been understood, and their accuracy have also been improved [Kataoka (2007), Tada (2008)]. Therefore, it is interesting to compare these SIMS data with the calibrated MC simulation. Figure 6 compares the SIMS B and As profiles and MC simulation. The SIMS B and As profiles near the peak region agree well with the MC data. Therefore, the MC simulation calibrated in the energy range of around few 10 keV can also predict the profiles in the energy range of around 1 keV.

Figure 7 also compares SIMS and MC results with combined  $S_e$  model of Eq. 8. We applied  $\theta$  of 1.65 to the other energies and ions and obtained good agreement. We therefore can predict ion implantation profiles over the energy region from 0.5 keV to more than 2000 keV with our MC.

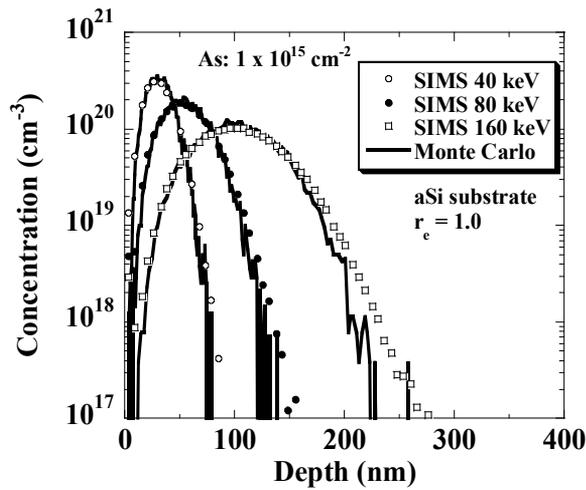


Fig. 5. Comparison of SIMS As profiles with MC with default  $r_e$  of 1.0.

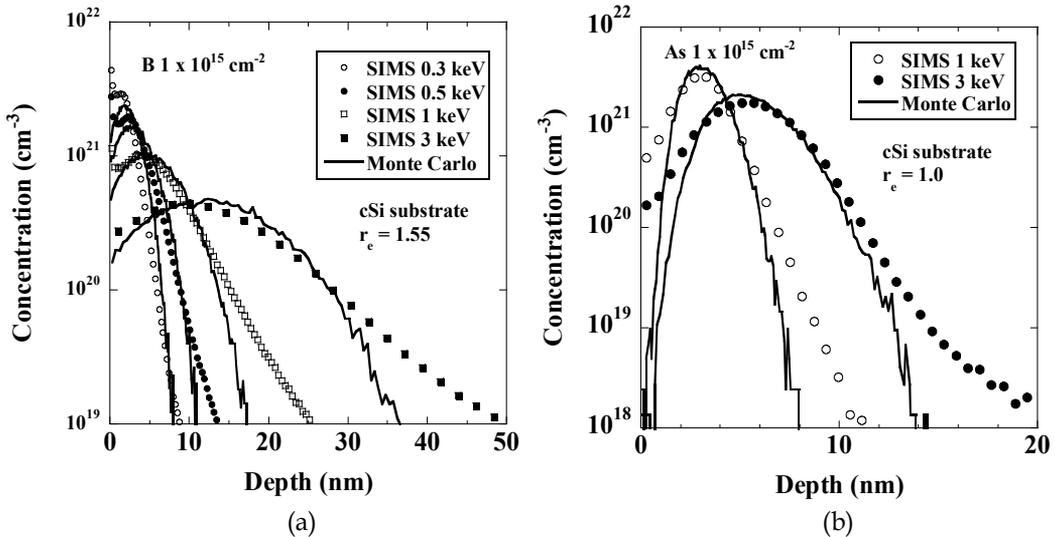


Fig. 6. Comparison of low-energy B and As SIMS profiles with MC.

Ziegler utilized the linear response method and treated the transition from Lindhard's model to Bethe's model more universally [Ziegler (2008)]. Although the Ziegler's model is physical one, we cannot obtain a good agreement as it is [Suzuki (2009)] and tune its parameters, which is not easy to handle.

As we pointed out in the Section 3, the Lindhard's  $S_e$  model becomes invalid in high-energy region, especially for B in the practical high-energy region of around MeV. Figure 7 compares B SIMS data with the MC simulation using Lindhard's  $S_e$  model. SIMS and MC results agree well at 400 keV. However, Lindhard's model predicts much shallower B profiles at 1200 and 2000 keV.

Figure 8 shows the dependence of stopping powers on energy. We need not care about this subject of the limitation of Lindhard's  $S_e$  model for P and As since  $eE_r$  values are much

larger than 5 MeV for P and As, as shown in Fig. 8. However, if we use much higher energies for these ions or light ions such as B, we should find optimal values of  $\theta$  for each combination of ion and substrate and may be  $r_{eff}$  since the model is empirical.

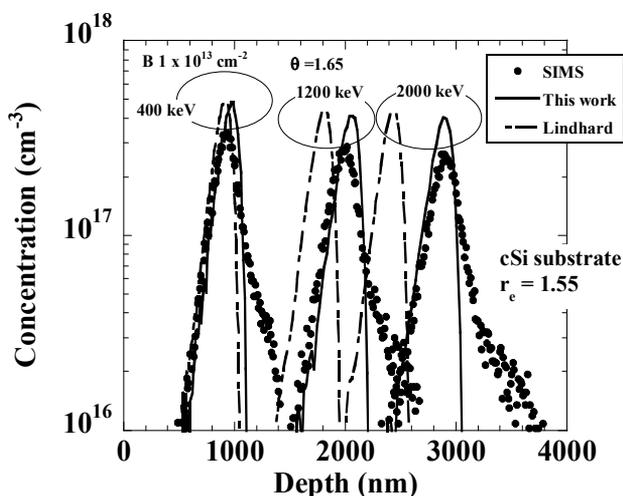


Fig. 7. Comparison of high-energy B SIMS data with MC using Lindhard and combined Se models.

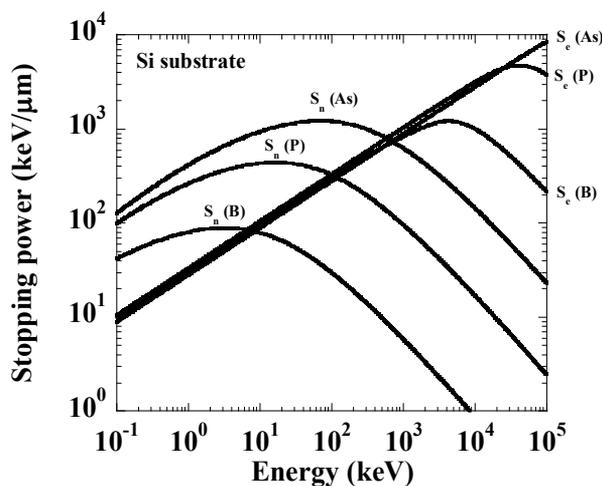


Fig. 8. Stopping powers of B, P, and As in Si substrate.

### 5. Database based on Pearson function

MC results are scattered in the low concentration region, and hence analytical fitting function such as Pearson IV is sometimes used instead of the MC raw data. Furthermore, if we convert the results to an analytical function, we can establish the database based on the parameters of the function and generate profiles for any ion implantation conditions by

interpolating the parameter values. In these cases, it is important how the analytical function can reproduce the MC results. Pearson function is vital, and hence it is a standard function to express ion implantation profiles. The parameters of the function can be evaluated from first four moments of the MC results. We show here that the simple use of the raw moments of the MC results sometimes induce significant error and how to solve it. Pearson IV is frequently predominately used among the Pearson function family. There exists long-standing discussion on validity of the selective use of Pearson IV [Ashworth (1990)]. We give some insight on it.

### 5.1 Pearson function family

Here, we briefly explain Pearson function family [Ashworth (1990), Selberherr (1984)] and show the definition of moments of the profiles and the relation of the moments to the Pearson function parameters.

We first convert the raw profile  $N(x)$  to the normalized one  $h(x)$  with respect to the dose for simple treatment, that is,

$$h(x) = \frac{N(x)}{\int_{x_a}^{x_b} N(x) dx} \quad (10)$$

Although the definition region is infinite plane for Pearson IV function, it is limited for some Pearson functions, and we hence describe the definition region as  $[x_a, x_b]$

The first moment parameter of projected range  $R_p$  is defined as

$$R_p = \int_{x_a}^{x_b} xh(x) dx \quad (11)$$

We convert the depth  $x$  with respect to  $R_p$  to  $s$  as

$$s = x - R_p \quad (12)$$

We also define  $s_a \equiv x_a - R_p, s_b \equiv x_b - R_p$ .

The  $n$ -th moment  $\mu_n$  can be evaluated as

$$\mu_n = \int_{s_a}^{s_b} s^n h(s) ds \quad (13)$$

Instead of the moments of  $\mu_2, \mu_3, \mu_4$  defined by Eq. 13, the following related parameters are used.

$$\Delta R_p = \sqrt{\mu_2} \quad (14)$$

$$\gamma = \frac{\mu_3}{\Delta R_p^3} \quad (15)$$

$$\beta = \frac{\mu_4}{\Delta R_p^4} \quad (16)$$

$\Delta R_p$  is called as straggling;  $\gamma$ , as skewness; and  $\beta$ , as kurtosis. Pearson function family is generated from the differential equation

$$\frac{dh}{ds} = \frac{(s-a)h}{b_0 + b_1s + b_2s^2} \quad (17)$$

Modifying Eq. 17 to

$$(b_0 + b_1s + b_2s^2) \frac{dh}{ds} = (s-a)h \quad (18)$$

multiplying  $s^n$ , and integrating it in the region of  $[s_a, s_b]$ , we obtain

$$\int_{s_a}^{s_b} (b_0s^n + b_1s^{n+1} + b_2s^{n+2}) \frac{dh}{ds} ds = \int_{s_a}^{s_b} (s^{n+1} - as^n) h ds \quad (19)$$

We then obtain

$$\left[ (b_0s^n + b_1s^{n+1} + b_2s^{n+2}) h \right]_{s_a}^{s_b} - \int_{s_a}^{s_b} (nb_0s^{n-1} + (n+1)b_1s^n + (n+2)b_2s^{n+1}) h ds = \int_{s_a}^{s_b} (s^{n+1} - as^n) h ds \quad (20)$$

Imposing

$$\lim_{s \rightarrow s_a, s_b} [s^{n+2}h] = 0 \quad (21)$$

we obtain

$$nb_0\mu_{n-1} + [(n+1)b_1 - a]\mu_n + [(n+2)b_2 + 1]\mu_{n+1} = 0 \quad (22)$$

Substituting  $n = 0, 1, 2, 3$  in Eq. 22, and utilizing  $\mu_0 = 1, \mu_1 = 0$ , each parameter in Eq. 17 can be expressed with the moment parameters as

$$b_0 = -\frac{4\beta - 3\gamma^2}{A} \Delta R_p^2 \quad (23)$$

$$b_1 = -\frac{\beta + 3}{A} \gamma \Delta R_p \quad (24)$$

$$b_2 = -\frac{2\beta - 3\gamma^2 - 6}{A} \quad (25)$$

$$a = b_1, \quad (26)$$

where

$$A = 10\beta - 12\gamma^2 - 18 \quad (27)$$

The definition region of  $[s_a, s_b]$  is related to the singular points of the denominator of Eq. 17. When the denominator is first order equation with respect to  $s$ , that is  $b_2 = 0$ , the related singular point  $\xi_s$  is

$$\xi_s = -\frac{b_0}{b_1} \quad (28)$$

When the denominator is second order equation with respect to  $s$ , that is  $b_2 \neq 0$ , the related singular points  $\xi_1, \xi_2$  are

$$\xi_1 = \frac{-b_1 - \sqrt{b_1^2 - 4b_0b_2}}{2b_2} \quad (29)$$

$$\xi_2 = \frac{-b_1 + \sqrt{b_1^2 - 4b_0b_2}}{2b_2} \quad (30)$$

When the denominator has double roots, the related singular point  $\xi_D$  is

$$\xi_D = -\frac{b_1}{2b_2} \quad (31)$$

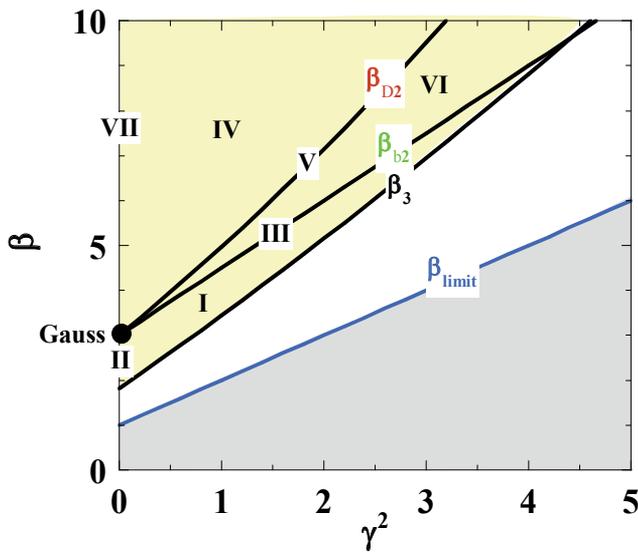


Fig. 9. Domains for each Pearson function. Any function that is always positive, concentration should be  $\beta > \beta_{limit}$

Figure 11 shows the domains for various Pearson functions in  $(\gamma^2, \beta)$  plane, where

$$\beta_{D2} = \frac{3(13\gamma^2 + 16) + 6(\gamma^2 + 4)^{\frac{3}{2}}}{32 - \gamma^2} \quad (32)$$

$$\beta_{b2} = \frac{3}{2}\gamma^2 + 3 \quad (33)$$

$$\beta_3 = \frac{9 \left[ 6\gamma^2 + 5 + \sqrt{\frac{9}{16}\gamma^6 + 8\gamma^4 + 25(\gamma^2 + 1)} \right]}{50 - \gamma^2} \quad (34)$$

$$\beta_{limit} = \gamma^2 + 1 \quad (35)$$

$\beta_3$  is the limit for Pearson function, and  $\beta_{limit}$  is the limitation for general distribution function that holds positive over entire definition region [Ashworth (1990)].

The corresponding  $(\gamma^2, \beta)$  region, definition region  $[s_a, s_b]$ , and analytical form of each Pearson function are as follows [Ashworth (1990)].

**Gauss:**  $\gamma = 0, \beta = 3; [-\infty, \infty]$

$$\ln h(s) = \frac{s^2}{2b_0} + \ln K, \quad (36)$$

where  $K$  is the factor to realize normalization, and it is also used for the other functions.

**Pearson IV, VII:**  $\beta > \beta_{D2}; [-\infty, \infty]$

The expression for Pearson IV is given by

$$\ln h(s) = \frac{1}{2b_2} \ln |b_0 + b_1s + b_2s^2| - \frac{2b_1 + \frac{b_1}{b_2}}{\sqrt{4b_0b_2 - b_1^2}} \tan^{-1} \left( \frac{2b_2s^2 + b_1}{\sqrt{4b_0b_2 - b_1^2}} \right) + \ln K \quad (37)$$

Pearson VII is the special case of Pearson IV, where  $\gamma = 0$ , that is, identical to  $b_1 = 0$  in Eq. 37 and hence the expression for Pearson VII is given by

$$\ln h(s) = \frac{1}{2b_2} \ln |b_0 + b_2s^2| + \ln K \quad (38)$$

**Pearson V:**  $\beta = \beta_{D2}; [-\infty, \xi_D]$  for  $\gamma < 0, [\xi_D, \infty]$  for  $\gamma > 0$

$$\ln h(s) = \frac{1}{2b_2} \ln |b_0 + b_1s + b_2s^2| + \frac{2b_1 + \frac{b_1}{b_2}}{2b_2s + b_1} + \ln K \quad (39)$$

**Pearson VI:**  $\beta_{b2} < \beta < \beta_{D2}; [-\infty, \xi_1]$  for  $\gamma < 0, [\xi_2, \infty]$  for  $\gamma > 0$

$$\ln h(s) = \frac{1}{2b_2} \ln |b_0 + b_1s + b_2s^2| - \frac{b_1 + \frac{b_1}{2b_2}}{\sqrt{4b_0b_2 - b_1^2}} \ln \left| \frac{2b_2 + b_1 - \sqrt{b_1^2 - 4b_0b_2}}{2b_2 + b_1 + \sqrt{b_1^2 - 4b_0b_2}} \right| + \ln K \quad (40)$$

**Pearson I, II:**  $\beta_3 < \beta < \beta_{b2}; [\xi_1, \xi_2]$

The expression of Pearson I is the same as that for Pearson VI given by

$$\ln h(s) = \frac{1}{2b_2} \ln |b_0 + b_1s + b_2s^2| - \frac{b_1 + \frac{b_1}{2b_2}}{\sqrt{4b_0b_2 - b_1^2}} \ln \left| \frac{2b_2 + b_1 - \sqrt{b_1^2 - 4b_0b_2}}{2b_2 + b_1 + \sqrt{b_1^2 - 4b_0b_2}} \right| + \ln K \quad (41)$$

Pearson I is the special case of Pearson I, where  $\gamma = 0$ , that is  $b_1 = 0$  in Eq. 41, and is given by

$$\ln h(s) = \frac{1}{2b_2} \ln |b_0 + b_2 s^2| + \ln K \quad (42)$$

**Pearson III:**  $\beta = \beta_{b_2}; [-\infty, \xi_s]$  for  $\gamma < 0, [\xi_s, \infty]$  for  $\gamma > 0$

$$\ln h(s) = \frac{1}{b_1} s - \left(1 + \frac{b_0}{b_1^2}\right) \ln \left|s + \frac{b_0}{b_1}\right| + \ln K \quad (43)$$

It seems that we can simply obtain the moments of MC results, and select the function and generate the corresponding Pearson function. However, there are some problems in this procedure, as shown in the following sections.

### 5.2 Monte Carlo tracing to the negative plane

When light impurities are ion-implanted into a substrate composed of heavy atoms, the backscattering becomes significant, and the number of them is not negligible. The resultant profiles are cut at the surface. On the other hand, the Pearson profile is continuous over the whole area. Therefore, the moments of this MC result cut at the surface may induce inaccurate Pearson distribution.

Figure 10 shows B profiles ion-implanted in W substrates. It clearly shows that the back scattering is significant and the profiles are cut at the surface. The dashed line corresponds to the Pearson profiles using the moments of the MC results. The agreement is not as good as expected.

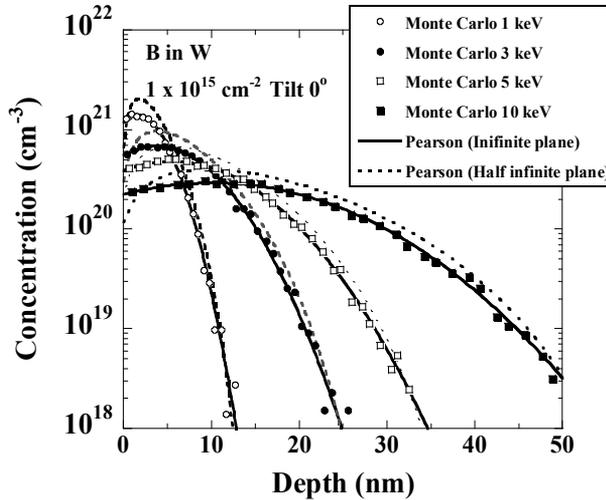


Fig. 10. MC simulation of B ion implantation profiles in W substrates. Pearson function using raw moment parameters, and moment parameters of MC tracing to the negative plane are also shown.

This situation occurs when the energy becomes quite low even in cases of B in Si substrate although the backscattering is not so significant for energy regions of around few 10 keV.

Figure 11 shows the B profiles ion implanted at 0.5 keV. It is also clear that the back scattering is significant in this case, and the Pearson profile deviates from the MC result as is also the case in W substrate.

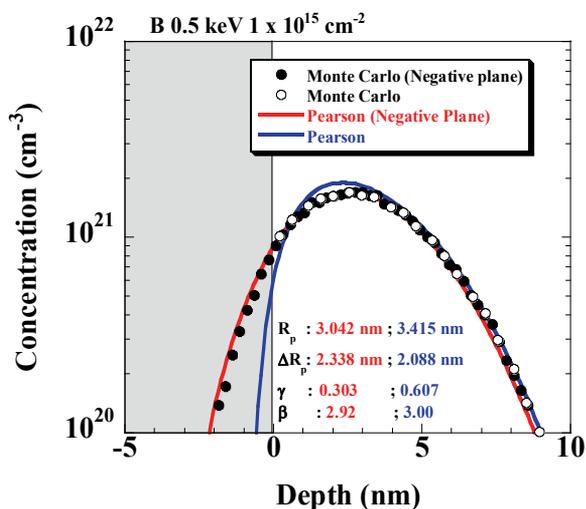


Fig. 11. MC simulation of B ion implantation profiles in Si substrates. Pearson function using raw moment parameters, and moment parameters of MC tracing to the negative plane are also shown.

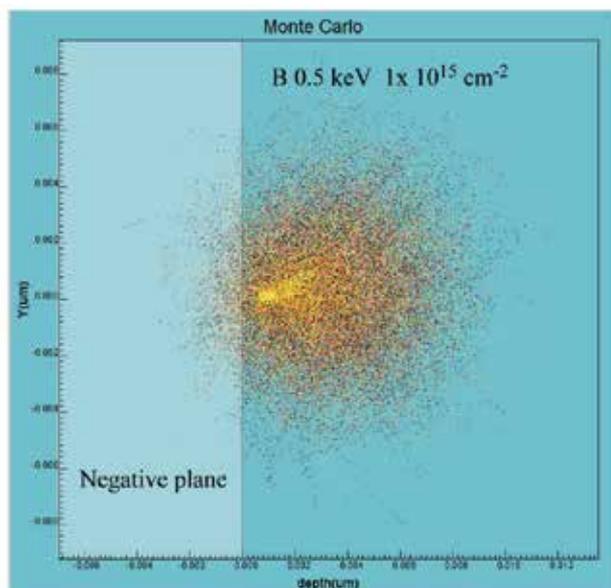


Fig. 12. MC tracing to the negative plane.

We propose virtually setting the substrate in a negative plane (infinite plane) and tracing ions that are backscattered at the surface or that have escaped from the bulk to the surface,

as shown in Fig. 12 [Suzuki (2010b)]. By extracting the moments from the results, we reproduced the MC results more accurately than in the case of the standard MC simulation moments, as shown in Figs. 10 and 11.

### 5.3 $\beta$ at high-energy region

We can overcome the problem of backscattering by using MC tracing to the negative plane, as shown in the previous section. We also show the other problem here, which is more fundamental one related to the limitation of Pearson function. This is also related to the availability to use Pearson IV function among Pearson function family.

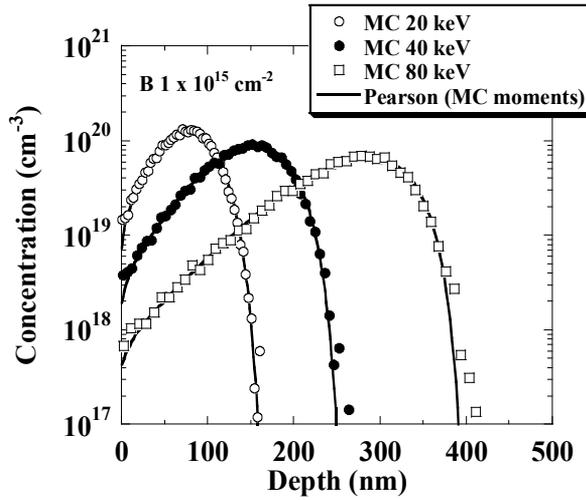


Fig. 13. MC simulation of B ion implantation profiles in Si substrates in the moderate-energy region between 20 and 80 keV. Pearson function using raw moment parameters, and moment parameters of MC tracing to the negative plane are almost the same in this energy region

Figure 13 compares MC results for B profiles ion implanted at 20, 40, and 80 keV with Pearson function using its moment parameters. They agree well and hence using raw data of moment parameters gives accurate Pearson function as well as the low-energy region, as shown in Fig 11. The moment parameter values are almost invariant, independent of the MC simulation mode of tracing or non-tracing in this energy region.

Figure 14(a) shows  $(\gamma^2, \beta)$  extracted from Monte Carlo data of ion implanted profiles in Si and Ge substrates with the energy region up to 320 keV. It is noteworthy that they are almost on the line of

$$\beta = 2.661 + 1.852\gamma^2 \quad (44)$$

and also that  $(\gamma^2, \beta)$  is always outside Pearson IV region.

Figure 14(b) shows the  $(\gamma^2, \beta)$  extending for B profiles the energy region up to 5000 keV. It is noteworthy that the simple extension of Eq. 44 to the higher  $\gamma^2$  region is invalid, and further  $(\gamma^2, \beta)$  breaks the limitation of Pearson function. Therefore, the simple use of moment parameters is obviously invalid in high-energy region.

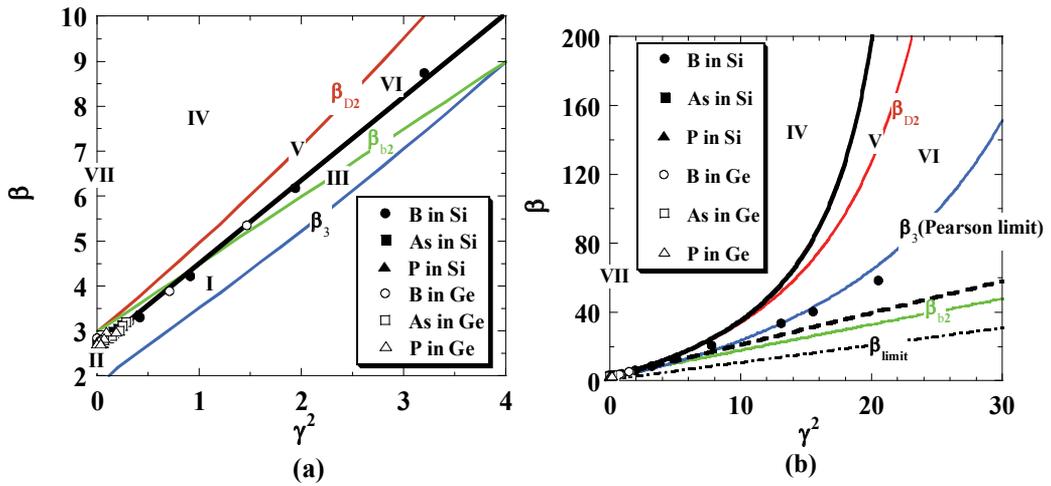


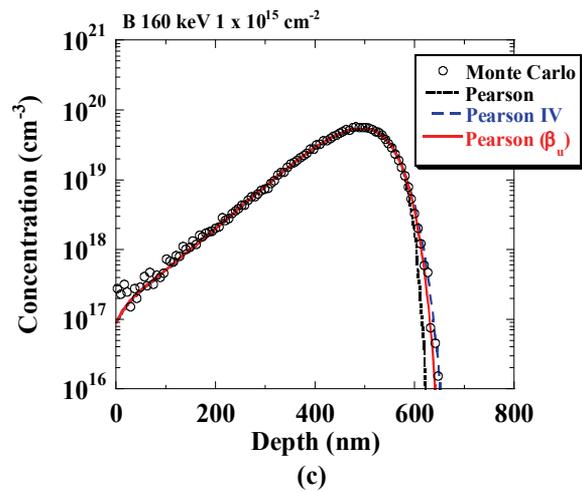
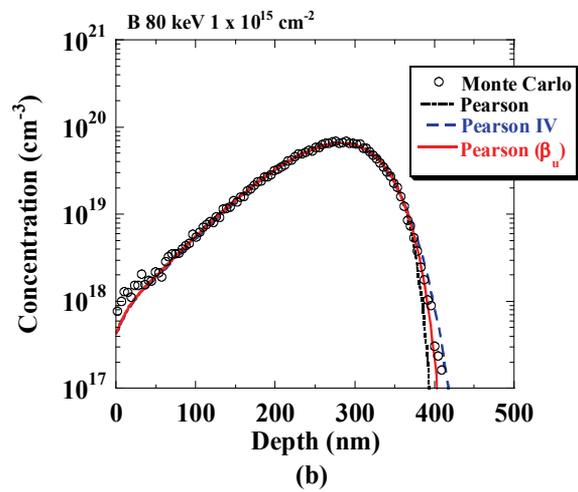
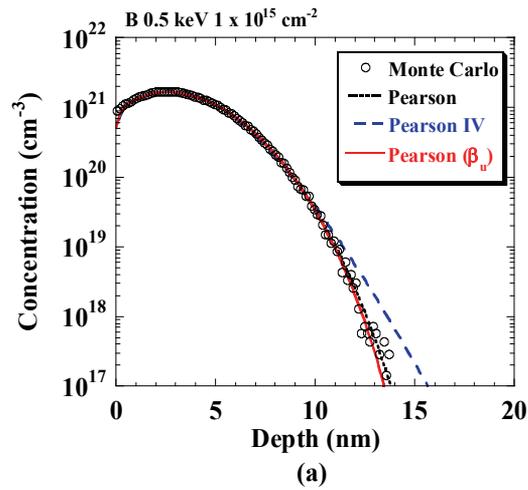
Fig. 14.  $\gamma^2 - \beta$  Relationship for various ion and Si and Ge substrates. (a) Energy region less than 360 keV. (b) Energy up to 5000 keV for B.

Based on the above data, we compare the MC data with Pearson using raw moment data, and Pearson IV with  $\beta$  forcing to  $\beta_{D2}$  of Eq. 32, as shown in Fig. 15.

The profile at an energy of 0.5 keV is symmetrical, that is,  $|\gamma|$  of the profile is small. The Pearson function using raw moment parameters effectively reproduces the MC data, while the Pearson IV profile deviates from the MC data in the low tail concentration region. The MC profile at 80 keV becomes asymmetrical, and the Pearson profile deviates from the MC data, and MC data is in between the Pearson and Pearson IV profiles. When we further increase the energy to 160 keV, the Pearson profile becomes inaccurate, and Pearson IV becomes better. Further, the Pearson profile clearly becomes inaccurate even in the peak region at 1000 keV, while Pearson IV readily reproduces the MC data. Consequently, the MC results can be well expressed by Pearson function using raw moment parameters when the profile is symmetrical and Pearson IV is inaccurate and vice versa when the profile is asymmetrical.

We think that the above information can be appreciated with the following.

The moments of the profile can be defined for any order, as shown in Eq. 13, while Pearson function uses only the first four moments. When a profile is rather symmetrical, it can be accurately expressed using the first four moment parameters, and the direct use of the moment values can generate the profile accurately, and different moment parameters such as those for Pearson IV induce inaccurate one. This is the case for low energy. When the profile is rather asymmetrical, the four moments are not enough to express the profile, that is, Pearson function is not available if we use original moment values. If we limit ourselves to use Pearson function, we should use moments different from the raw ones to improve the accuracy approximately. One way to modify the parameter is to increase  $\beta$  although we do not understand its mathematical reason. If we use Pearson IV in this case, we use larger  $\beta$  than the raw ones, which ensure better agreement. This is the case for high energy. Therefore, Pearson IV expresses the profile more accurately than the Pearson using the raw moment values in this case.



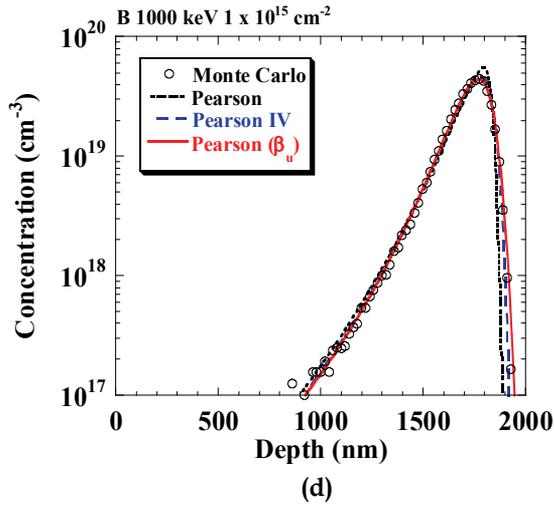


Fig. 15. MC simulation of B ion implantation profiles in Si substrates. Pearson function using raw moment parameters, and moment parameters of MC tracing to the negative plane are also shown. The Pearson using the proposed B is also shown. (a) 0.5 keV, (b) 80 keV, (c) 160 keV, and (d) 1000 keV.

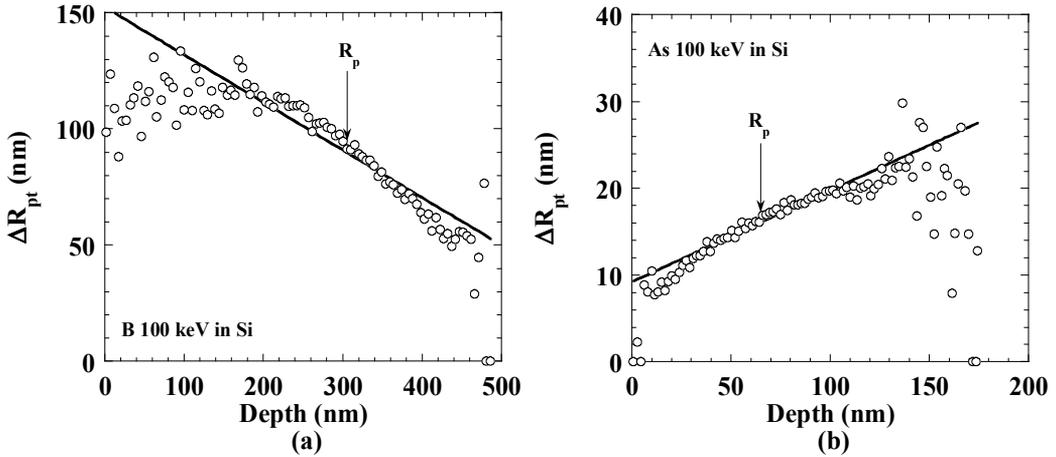


Fig. 16. Depth dependent lateral straggling.

Based on the above information, we propose to use  $\beta$  of

$$\beta_u = \frac{25(2.661 + 1.852\gamma^2)}{25 - \gamma^2} \quad (45)$$

We denote this  $\beta$  as  $\beta_u$ .  $\beta_u$  is shown in Fig. 14 as a solid line, and it is almost the same as Eq. 44 in low-energy ranges, as shown in Fig. 14(a), and it enters the Pearson IV domain for large  $\gamma^2$ , as shown in Fig. 14 (b). Figure 15 also shows the profiles using  $\beta_u$ , where any MC profile can be accurately expressed.

We can also evaluate lateral straggling  $\Delta R_{pt}$  from the MC results by integrating over the lateral direction. It is also pointed out that the lateral straggling depends on depth [Suzuki (2001)]. Figure 16 shows evaluated lateral straggling of B and As ion implantation.  $\Delta R_{pt}$  decreases with depth for B, and increases for As. It is clear that the lateral straggling depends on the depth linearly near the depth of  $R_p$ . Therefore, we evaluate the lateral straggling as

$$\Delta R_{pt} = \Delta R_{pt0} + m(x - R_p), \quad (46)$$

where  $\Delta R_{pt0}$  is the  $\Delta R_{pt}$  at  $R_p$ . Using the form of Eq. 46, we can extract  $m$  from the MC results. Figure 16 also shows the extracted  $\Delta R_{pt}$  as solid lines.

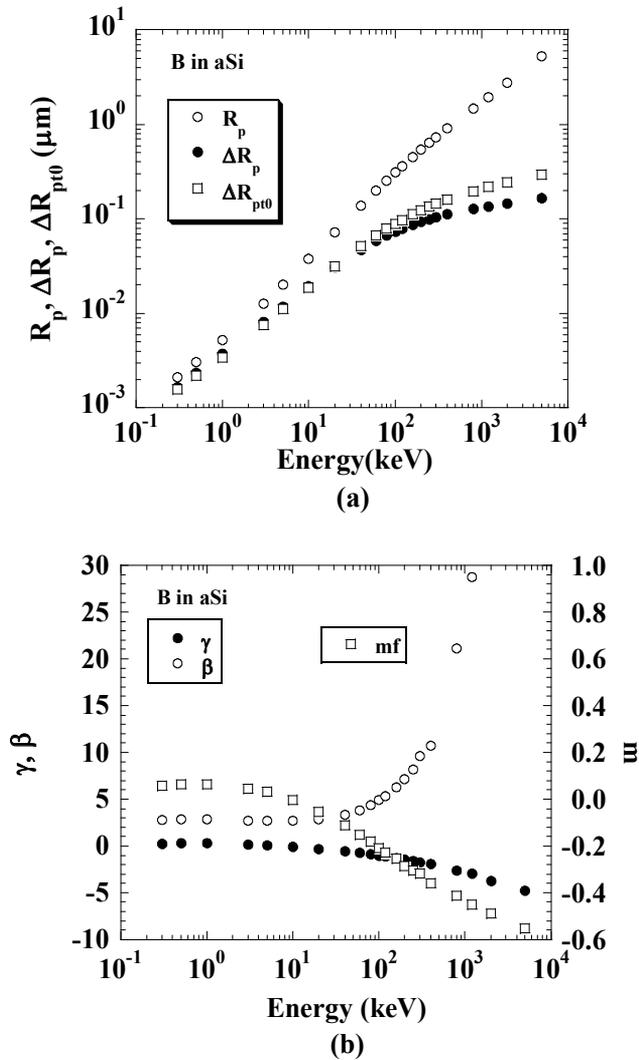


Fig. 17. Moment parameters of B ion implantation profiles.

Using the above procedure of MC tracing to the negative plane and  $\beta_u$ , we can establish a database of B, P, As, In, and Sb In ion implantation, as shown in Figs. 17, 18, 19, 20, 21, respectively. Using the database, we can generate ion implantation profiles instantaneously for any ion implantation conditions.

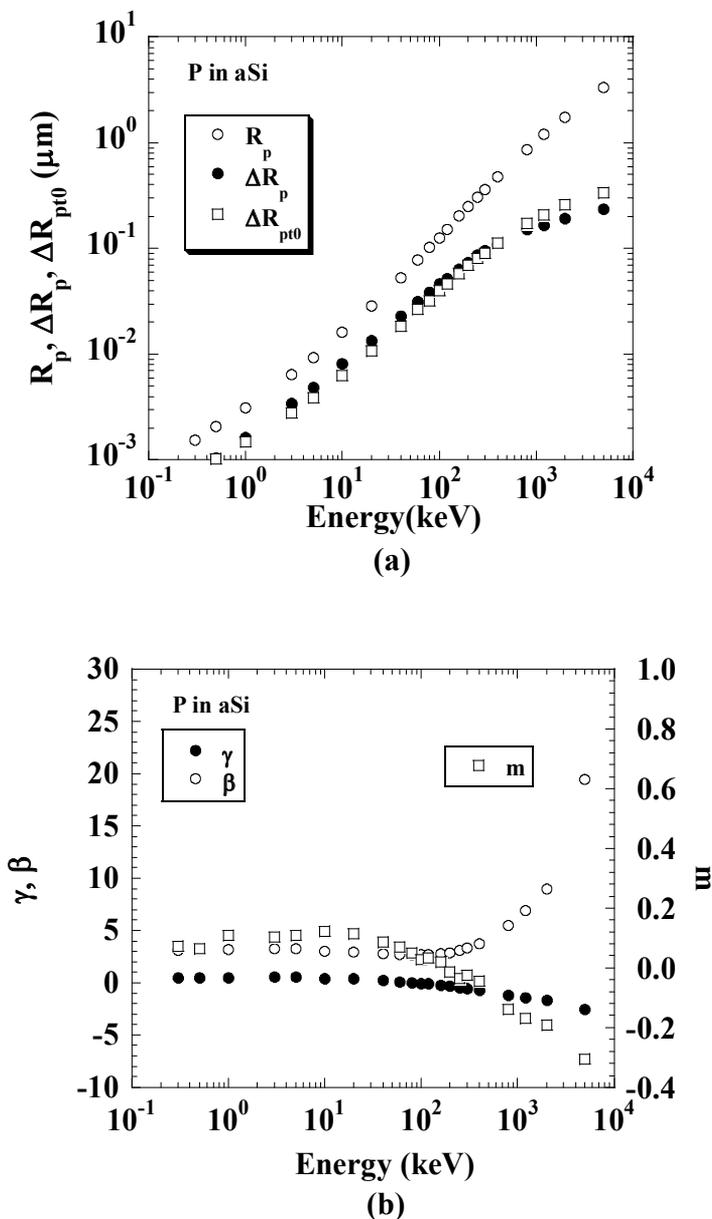


Fig. 18. Moment parameters of P ion implantation profiles.

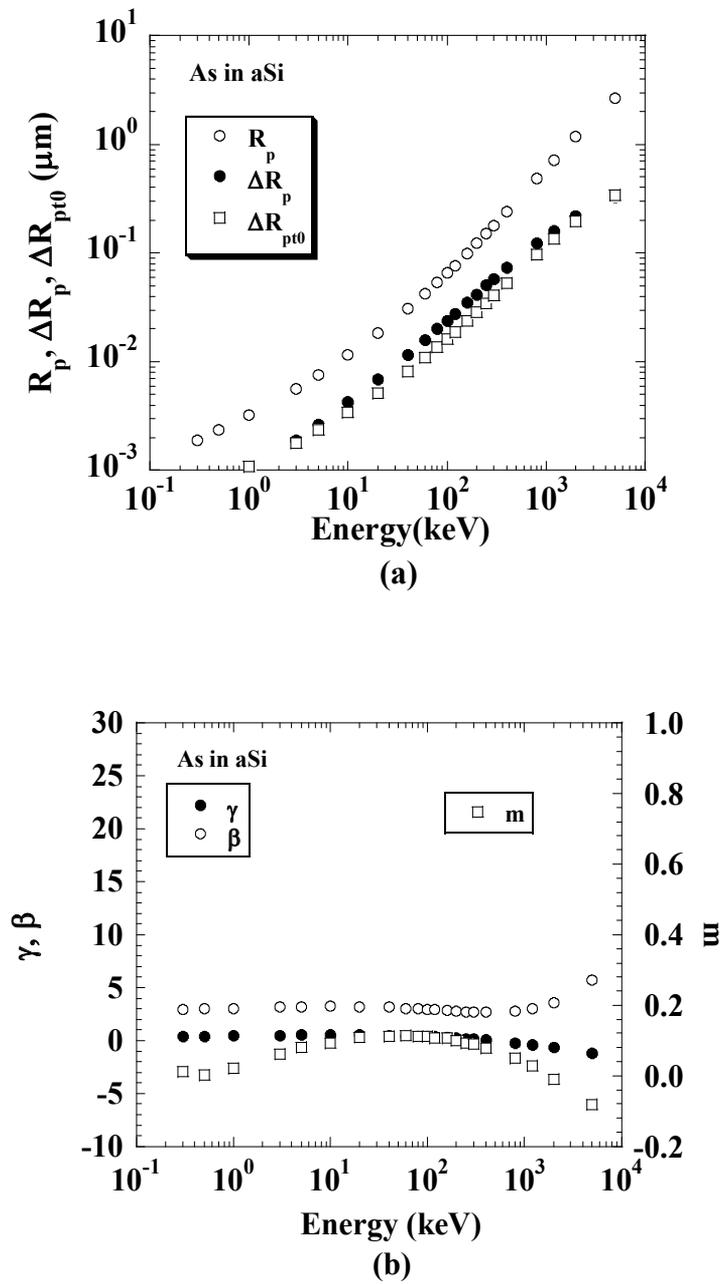


Fig. 19. Moment parameters of As ion implantation profiles.

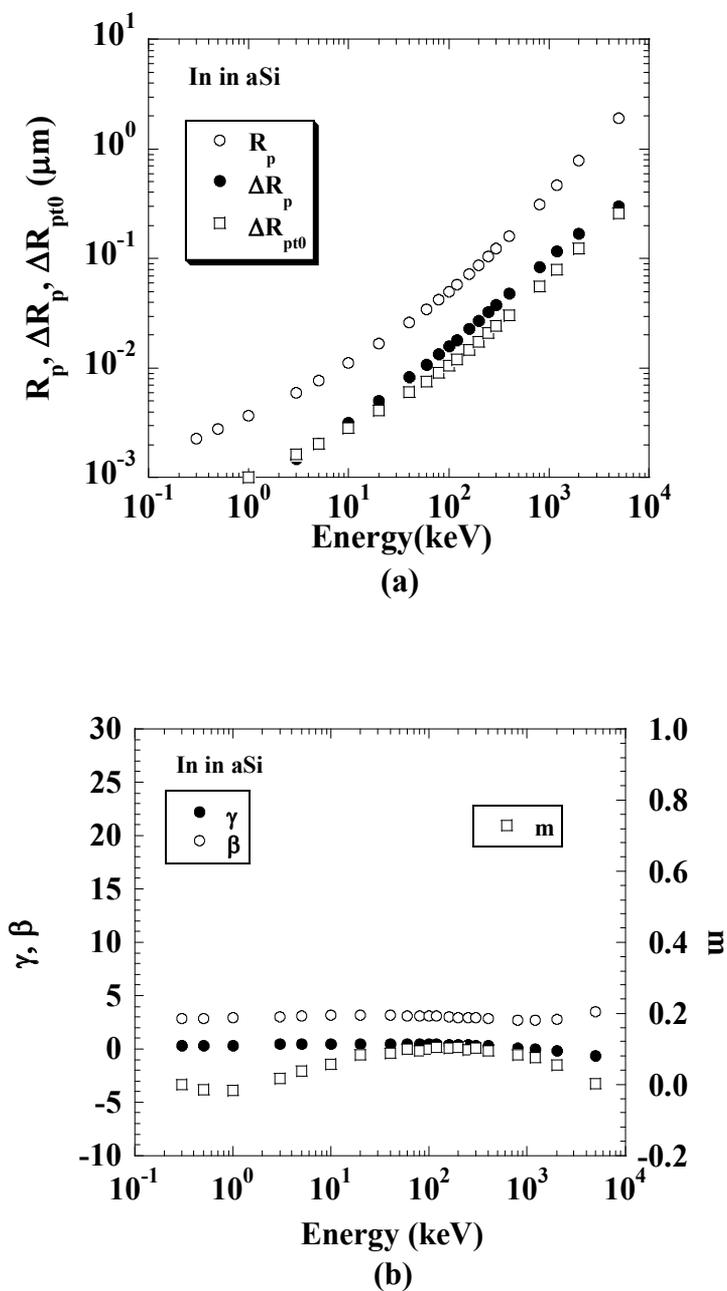


Fig. 20. Moment parameters of In ion implantation profiles.

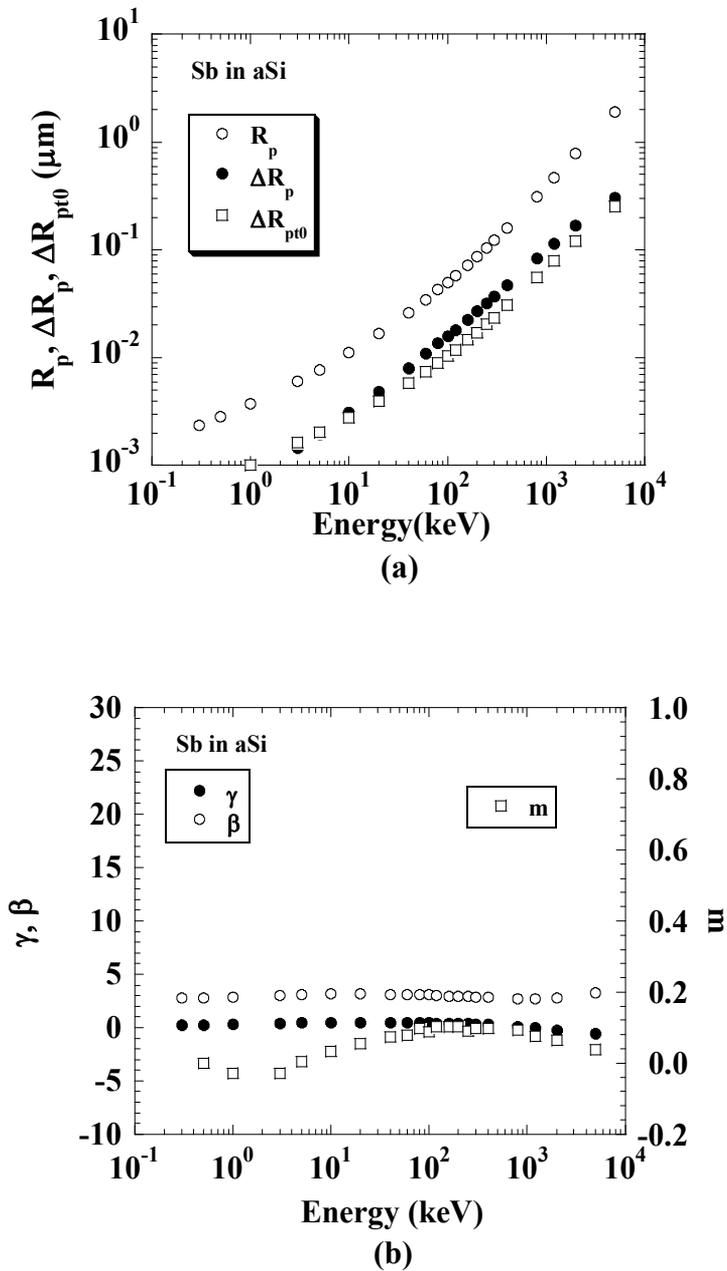


Fig. 21. Moment parameters of Sb ion implantation profiles.

## 6. Comparison of MC with TEM data

Figure 22 shows the cross-sectional TEM image of the amorphous layer formed by Ge ion implantation at an energy of 10 keV and various doses. We do not observe continuous amorphous layer at a dose of  $10^{13} \text{ cm}^{-2}$ , continuous amorphous layer with a thickness  $d_a$  of 9.4 nm at a dose of  $10^{14} \text{ cm}^{-2}$  with non-clear amorphous/crystal (a/c) interface, larger  $d_a$  of 20.2 nm with clear a/c interface at a dose of  $10^{15} \text{ cm}^{-2}$ , and further larger  $d_a$  of 24.5 nm with clear a/c interface at a dose of  $5 \times 10^{15} \text{ cm}^{-2}$ .

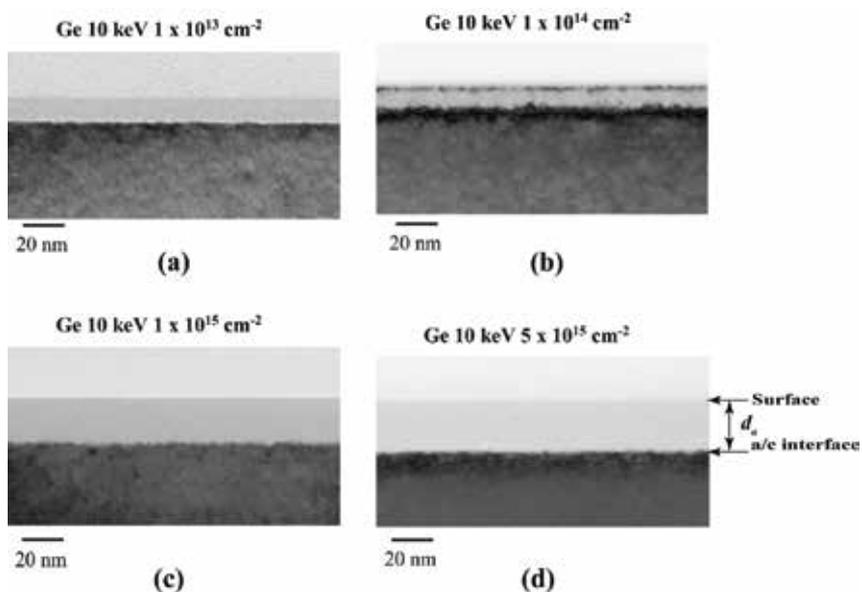


Fig. 22. Cross-sectional TEM image of the amorphous layer formed by Ge ion implantation.

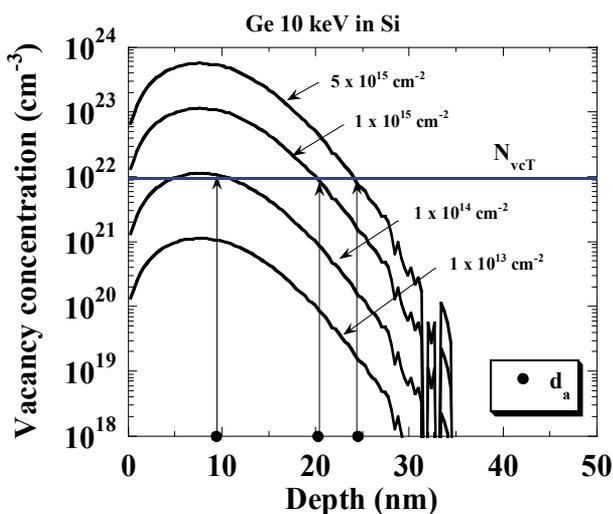


Fig. 23. Vacancy concentration induced by Ge ion implantation.

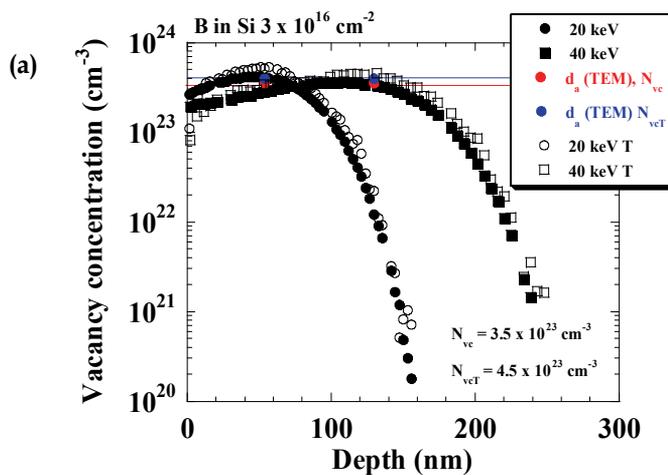
MC can evaluate accurately the transferred energy and vacancy concentration, as shown in Section 3. It should be noted that MC cannot predict absolute vacancy concentration since it does not consider temperature and hence the recombination of generated vacancy and recoiled substrate atoms. However, MC results neglecting the recombination may be able to be related to the amorphous layer thickness empirically.

Figure 23 shows the vacancy concentration evaluated by MC with the tracing mode. The amorphous layer thicknesses evaluated from Fig. 22 are also shown. It is noteworthy that the vacancy concentration at depth of  $d_a$  is almost the same, independent of dose. Therefore, we may relate  $d_a$  to the critical vacancy concentration denoted by  $N_{vcT}$ . This also well explains that the continuous amorphous layer is not formed with a dose of  $10^{13} \text{ cm}^{-2}$  since the peak vacancy concentration is lower than  $N_{vcT}$ .

It is also interesting that  $d_a$  is near the peak region at a dose of  $10^{14} \text{ cm}^{-2}$  where is the vacancy concentration is still high ever deeper than  $d_a$ . On the other hand, the gradient of vacancy concentration is high for the doses of  $1 \times 10^{15} \text{ cm}^{-2}$  and  $5 \times 10^{15} \text{ cm}^{-2}$ . This means that the vacancy concentration decreases drastically in the deeper region than  $d_a$ . Therefore, the clearness of the a/c interface can be related to the gradient of the vacancy concentration at  $d_a$ .

Figure 24 shows the comparison of the amorphous layer thickness with the vacancy concentration for various ions and energies. It is noteworthy that the vacancy concentration at a depth of  $d_a$  is almost the same, independent of energy, as shown in Fig. 23, but it depends on ions. In the MC simulation, the two types of MC calculation mode of tracing and non-tracing are shown. When we trace the recoiled substrate atom, we denote it as T.

There is no significant difference between tracing and non-tracing mode when the ion is light as B, but the difference becomes significant when the ion is heavy as As. The trajectories of B and As implanted in Si substrate are shown in Fig. 25 as black dotted lines, and recoiled Si trajectories are shown as yellow dotted lines. The recoiled Si does not go far from the ion trajectory path of B, and the distribution of vacancy for both modes are almost the same. While the recoiled Si goes far from the As trajectory path and generates much more recoiled Si, the profile of recoiled Si distribution is significantly different from the ion trajectory site. This is the reason why the vacancy distributions shows difference in the two modes.



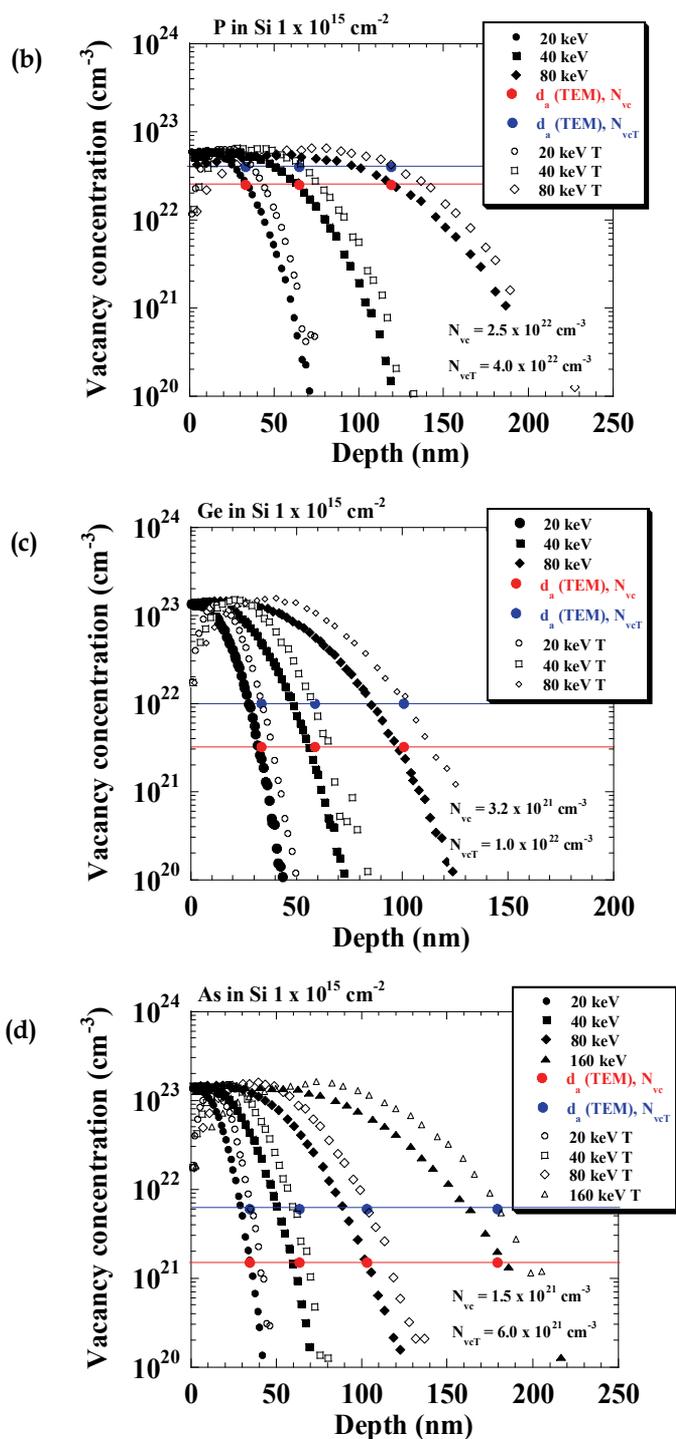


Fig. 24. Vacancy concentration with two modes: Tracing and non-tracing modes. (a) B in Si, (b) P in Si, (c) Ge in Si, (d) As in Si.

Therefore, the critical concentration of vacancy at the a/c interface is different although both are rather independent of energy. It is noteworthy that the vacancy concentration at the a/c interface is almost constant for each trace mode although the value is different in some cases. Therefore, we can predict  $d_a$  by evaluating the vacancy concentration in the MC simulation with both modes although the tracing mode is more physical.

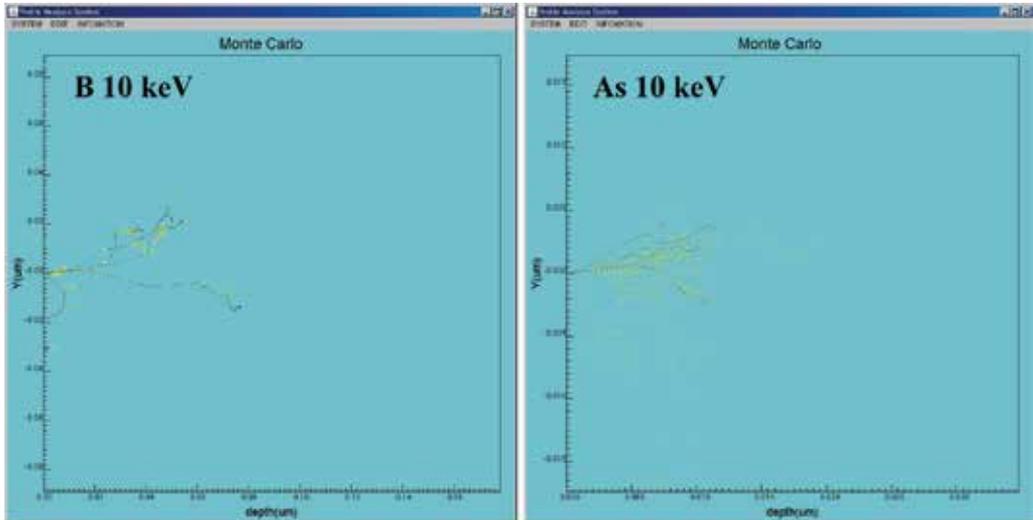


Fig. 25. MC simulation with tracing mode. B and As trajectories are shown. Black dotted lines correspond to the ion trajectories, and yellow ones to recoiled Si.

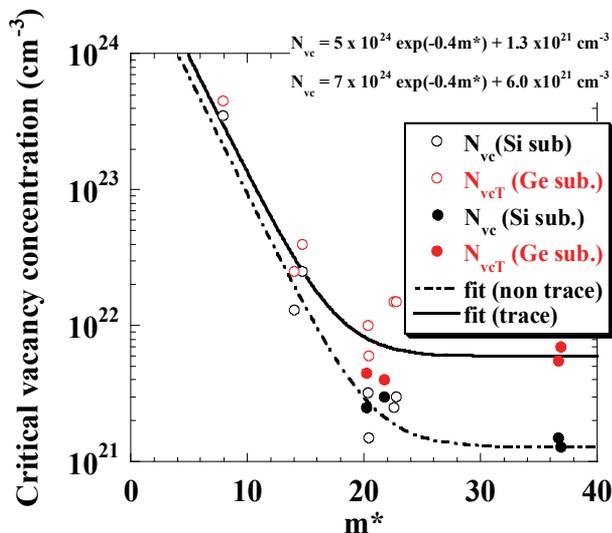


Fig. 26. Dependence of critical vacancy concentration on reduced mass.

We hope to predict  $d_a$  for an unknown incident ion and substrate system. Fig. 26 shows the dependence of critical vacancy concentration on reduced mass. Any point is almost on lines of

$$N_{vc} = 5 \times 10^{24} \exp(-0.4m^*) + 1.3 \times 10^{21} \text{ cm}^{-3} \quad (47)$$

$$N_{vcT} = 7 \times 10^{24} \exp(-0.4m^*) + 6.0 \times 10^{21} \text{ cm}^{-3}, \quad (48)$$

where  $m^*$  is the reduced mass defined by

$$\frac{1}{m^*} = \frac{1}{m_1} + \frac{1}{m_{sub}}, \quad (49)$$

although the physical reason is not clear, we can use Eqs. 47 and 48 to predict  $d_a$  as the initial guess of amorphous layer thickness by MC.

Prussin also analyzed the amorphous layer thickness based on Brices's energy deposition model. They also used critical deposition energy like  $N_{vc}$  or  $N_{vcT}$  in Cerva and our analysis. Cerva used the same  $N_{vcT}$  for P and As in Si substrate. However, we used a different one, and it is also different in ref. [Prussin (1984)]. The dependence becomes more clear when we used a light ion such as B whose  $N_{vc}$  or  $N_{vcT}$  is quite high in our analysis and also in Prussin (1984), while Cerva uses the constant  $N_{vcT}$ , independent of ions. We think that our result is plausible since B does not recoil Si atom so far from the original lattice site, and recombination aptly occurs through self annealing.

Prussin also expressed the amorphous layer thickness  $d_a$  with an empirical form of

$$d_a = R_p + n\Delta R_p, \quad (50)$$

where  $n$  is the fitting parameter and changes with ions and dose. We applied physical appreciation to Eq. 50 and proposed the modified one as

$$d_a = \begin{cases} R_p + \sqrt{2}\Delta R_p \operatorname{erfc}^{-1}\left(\frac{2\Phi_{\%}}{\Phi}\right) & \Phi \geq 2\Phi_{\%} \\ 0 & \Phi < 2\Phi_{\%} \end{cases}, \quad (51)$$

where  $\Phi$  is the dose and  $\Phi_{\%}$  is the through dose defined by

$$\Phi_{\%} = \int_{d_a}^{\infty} N(x) dx, \quad (52)$$

where  $N(x)$  is the ion concentration.  $\Phi_{\%}$  is regarded as constant if the ion and substrate are defined independent of energy and dose. If we assume Gaussian profile, we can perform the integration and obtain the analytical form as

$$\begin{aligned} \Phi_{\%} &= \int_{d_a}^{\infty} \frac{\Phi}{\sqrt{2\pi}\Delta R_p} \exp\left[-\left(\frac{x-R_p}{\sqrt{2}\Delta R_p}\right)^2\right] dx \\ &= \frac{\Phi}{2} \operatorname{erfc}\left(\frac{d_a-R_p}{\sqrt{2}\Delta R_p}\right) \end{aligned} \quad (53)$$

Let us combine this formula to MC. We then analyze Ge ion implantation. We can predict  $d_a = 59$  nm,  $R_p = 31.5$  nm, and  $\Delta R_p = 12.0$  nm for Ge ion implantation at 40 keV with a dose of  $1 \times 10^{15} \text{ cm}^{-2}$  from the MC simulation. Substituting these parameters to Eq. 53, we obtain  $\Phi_{\%c} = 1.1 \times 10^{13} \text{ cm}^{-2}$ . We then use this for any ion implantation condition of dose and energy. Figure 27 compares the experimental and theoretical amorphous layer thicknesses. We obtained a good agreement over the wide ion implantation condition. The agreement is rather bad for a dose of  $10^{14} \text{ cm}^{-2}$ . However, the a/c interface is not clear for this dose.

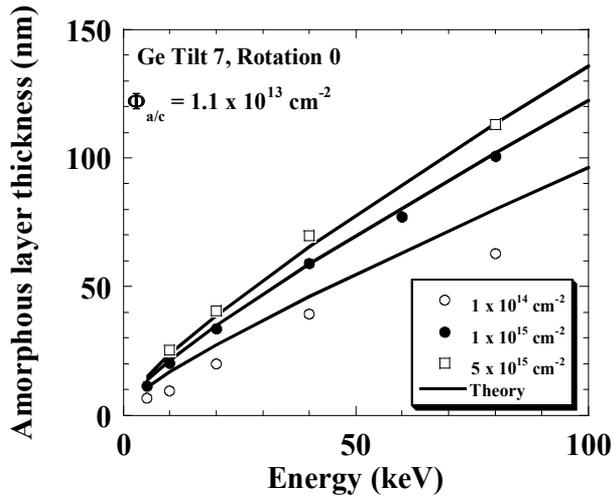


Fig. 27. Comparison of experimental  $d_a$  with the theoretical model using fixed  $\Phi_{\%c} = 1.1 \times 10^{13} \text{ cm}^{-2}$ .

In our analysis, we neglect the channeling phenomenon, which is a prominent feature for profiles in crystalline substrates. Although the channeling effect may be significant for the profiles in cSi, our successful analysis means that channeling is obviously negligible for the amorphization. This can be explained by the small fraction of channeling ions and by the fact that damage is primarily produced by non-channeling ions.

## 7. Conclusion

We showed that Monte Carlo simulation is vital to predict ion implantation profiles as well as amorphous layer thickness resulting from ion implantation. The MC results for ion implantation profiles deviate from the experimental data in some cases as it is. However, if we tune the parameter of electron stopping power for one energy, we can accurately predict the profiles for any energy. The ion implantation profiles for the high-energy region cannot be well reproduced by Lindhard electron stopping power model, but it can be reproduced by using the combined model of Lindhard and modified Bethe models. We also showed how to extract the parameters of MC data to generate the Pearson function. Simple use of moment parameters induces error in some cases, and we propose the MC tracing to the negative plane, and universal  $\beta$  instead of its raw value. We also showed that the forcing Pearson IV function is valid in the high-energy region where the profile is rather

asymmetrical, but it induces error in the low-energy region where the profile is rather symmetrical. We can simply predict the amorphous layer thickness by evaluating the vacancy concentration. The critical vacancy concentration depends on the calculation method: tracing or non-tracing mode, but the value is independent of energy for both modes. We proposed the empirical critical vacancy concentration, and  $d_a$  for any combination of incident ion and substrate atoms can be predicted using this. We can also evaluate a parameter of through dose  $\Phi_{\%}$  by MC. Using the  $\Phi_{\%}$  combined with the database for moment parameters of the profile based on the MC, we can predict  $d_a$  instantaneously without using MC afterward.

## 8. References

- Ashworth D. G., Oven R., and Mundin B. (1990), Appl. Phys. D., vol. 23, pp. 870-876.
- Bethe H. A. (1930), Ann. Phys. (Leipzig), vol. 5, pp. 325-400.
- Biersack J. P., and Hagmark L. G. (1980), Nuclear Inst. And Meth., vol. 174, pp. 257-269.
- Brice K. B. (1975), *Ion implantation range and energy deposition distribution*, IFI/Plenum Publishing Corporation, ISBN: 0-306-67401-7, London.
- Cerva H., and Hobler G. (1992), J. Electrochem. Soc., vol. 139, No. 12, pp. 3631-3638.
- Chui C. O., Ramanathan S., Triplett B. B., McIntyre P. C., and Saraswat K. C. (2002), IEEE Electron Device Lett., EDL-23, No. 8, pp. 473-475.
- Dalton P. and Turner J. E., (1968), Health Physics Pergamon Press, vol. 15, pp. 257-262.
- FabMeister-IM: <http://www.mizuho-ir.co.jp/english/solution/ion/index.html>
- Hobler G. and Selberherr S. (1988), IEEE Trans on Computer-Aided Desing, vol. 7, No.2, pp. 174-180, 1988.
- Hofker W. K. (1975), Philips Res. Rep. Suppl., vol. 8, pp. 1-121.
- Holmstrom E., Kuronen A., and Nordlund K. (2008), Physical Review B, vol 78, 045202.
- Hu J. C., Chatterjee A., Mehrotra M., Xu J., Shiao W. -T., and Rodder M. (2000), Proceedings of Symposium on VLSI Tech., pp. 188-189. Honolulu, June 13-15, 2000.
- Jin J.-Y., Liu J., Jeong U., and Mehta S. (2002), J. Vac. Sci. Technol. B, Vol. 20, pp. 422-426.
- Kataoka Y. and Itani T. (2007), Surf. Interface Anal., vol. 39, pp. 826-831.
- Kim Y. S., Shimamune Y., Fukuda M., Katakami A., Hatada A., Kawamura K., Ohta H., Sakuma T., Hayami Y., Morioka H., Ogura J., Minami T., Tamura N., Mori T., Kojima M., Sukegawa K., Hashimoto K., Miyajima M., Satoh S., and Sugii T. (2006), Proceedings of Electron Devices Meeting, pp. 622-625. ISBN: 1-4244-0438-x, San Francisco, December 11-13, 2006.
- Lindhard J., and Scharff M. (1961), Phys. Rev., Vol. 124, No. 1, pp. 128-130.
- Mirabera S., Impellizzeri G., Bruno E., Romano L., Grimaldi M. G., and Priolo F., Napolitani E., and Carnera A. (2005), Appl. Phys. Lett., vol. 86, 121905.
- Pawlak B. J., Vandervorst W., Smith A. J., Cowern N. E., Colombeau B., and Pages X. (2005), Appl. Phys. Lett., Vol. 86, 101913.
- Prussin S., Margolese D. I., and Tauber R. N. (1984), J. Appl. Phys. Vol. 57, No. 2, pp. 180-185. Sentaurus Process <http://www.synopsys.com/Tools/TCAD/ProcessSimulation/Pages/SentaurusProcess.aspx>
- Selberherr S. (1984), *Analytical and simulation of semiconductor device*, Springer-Verlag, ISBN: 3-211-81800-6, Wien.
- Shang H., Okorn-Schmidt H., Ott J., Kozlowski P., Steen S., Jones E. C., Wong H.-S. P., and Hanesch W. (2003), IEEE Electron Device Lett., EDL-24, No. 4, pp. 242-244.

- Schoerner R., Friedrichs P., Peter D., and Stephani D. (1999), IEEE Electron Device Lett., EDL-20, No. 5, pp. 241-244.
- Solmi S., Landi E., and Baruffaldi F. (1990), J. Appl. Phys., pp. 3250-3258.
- SRIM-2003: <http://www.srim.org/>
- Suzuki K., Sudo R., Tada Y., Tomotani M., Feudel T., and Fichtner W. (1998), Solid-State Electronic, vol. 42, pp. 1671-1678.
- Suzuki K., Sudo R., and Nagase M. (2001), IEEE Trans. Electron Devices, ED-48, pp. 2803-2807.
- Suzuki K. and Tashiro H. (2003), IEEE Trans. Electron Devices, ED-50, pp. 1753-1757.
- Suzuki K., Tashiro H., Narita K., and Kataoka Y. (2004), IEEE Trans. Electron Devices, ED-51, pp. 663-668.
- Suzuki K., Kawamura K., Kikuchi Y., and Kataoka Y. (2006), IEEE Trans. Electron Devices, vol. ED-53, NO. 5, pp. 1186-1192.
- Suzuki K., Tada Y., Kataoka Y., and Nagayama T. (2009), J. Semiconductor Technology and Science, vol.9, No. 1, pp. 67-74.
- Suzuki K. (2010a), Fujitsu Scientific & Technical Journal, vo. 46, No. 3, pp. 307-317.
- Suzuki K., Tada Y., Kataoka Y., and Kojima S. (2010b), Proceedings of 18th Ion Implantation Tech., P2-24, Kyoto, June 6-11, 2010.
- Tada Y., Suzuki K., and Kataoka Y. (2008), Applied Surface Science, vol. 255, pp. 1320-1322.
- Tian S. (2003), J. Appl. Phys., vol. 93, No. 10, pp. 5893-5904.
- Weber O., T. Irisawa, Numata T., Harada M., Taoka N., Yamashita Y., Yamamoto T., Sugiyama N., Takenaka M., and Takagi S. (2007), Proceedings of Electron Devices Meeting, pp. 719 - 722. ISBN: 1-4244-1507-1, Washington DC, December 10-12, 2007.
- Ziegler J. F., Biersack J. P., and Litmark U. (2008), *SRIM The Stopping and Ranges of Ions in Matter*, SRIM Co., ISBN: 0-9654207-1-x, USA.

# Application of Monte Carlo Simulation in Industrial Microbiological Exposure Assessment

Javier Collado<sup>1</sup>, Antonio Falcó<sup>2</sup>, Dolores Rodrigo<sup>1</sup>,  
Fernando Sampedro<sup>1</sup>, M. Consuelo Pina<sup>1</sup> and Antonio Martínez<sup>1</sup>

<sup>1</sup>*Instituto de Agroquímica y Tecnología de Alimentos (CSIC),*

<sup>2</sup>*Universidad Cardenal Herrera-CEU*

*Spain*

## 1. Introduction

### 1.1 Food safety systems: HACCP and risk analysis

Nowadays, modern societies pay great attention to food safety. For this reason, tools such as Hazard Analysis and Critical Control Points (HACCP) and Risk Analysis (RA) have been developed, helping the production of safer foodstuffs and, consequently, reducing the number of foodborne illnesses. Nevertheless, there are certain differences between HACCP and RA, HACCP system is the main tool in evaluating and controlling foodborne hazards in food industries while risk analysis demonstrate to be an effective tool in designing, developing and evaluating control measures protecting public health inside a country or region. In addition, RA has been used for seeking solutions in commercial litigation as an objective tool to assess the risk that a food has to the consumer (Zwietering & Nauta, 2007). HACCP systems do however have some limitations, such as the inability to be linked with public health goals and the incapability to deal with the inherent variability in food systems due to its qualitative nature (Buchanan & Whiting, 1998; Hoornstra et al. 2001).

Risk analysis is a complex process consisting of three interconnected components: risk assessment (scientific component), risk management (legal component) and risk communication. Risk assessment consists of four steps, *hazard identification* where the hazard is related to the onset of illness by using data from foodborne outbreaks caused by food contaminated with the hazard, as well as its taxonomy and virulence factors associated with the development of the disease. Having identified the hazard, a *hazard characterisation* is carried out which identifies the aetiology of the disease caused by the biological agent, the influence of different subgroups on the disease symptoms and the dose-response relationship. The third step is the *exposure assessment* which deals with the ecology of the biological hazard, the critical steps of the production chain in which the hazard could be present or proliferate, and the control measures which determine the influence of environmental factors and process conditions on the survival of the biological agent. Finally, all the information gathered in the previous sections is used to produce an estimation of the

risk, known as *risk characterisation*, expressed as the probability that a population acquires the disease by the consumption of the food contaminated with the hazard.

Risk assessment has a qualitative or quantitative nature depending on the availability of data. The qualitative assessment is the most widely used because of the lack of data on consumption pattern, dose-response models, initial contamination, and survival of the microorganism after treatment and until the time of consumption. In this case the magnitude of the risk can be described as insignificant, low, medium and high.

The quantitative microbial risk assessment (QMRA) is more complex and it is based on the availability of specific quantitative data concerning the prevalence of the hazard in the product under study at different steps of the process chain, as well as, the necessary dose to produce a response in the host (dose-response relationship) and the use of mathematical models to characterise that response.

The limitations of HACCP commented earlier can be overcome by combine it with risk assessment providing to the HACCP system with quantitative data for CCP establishment. Nevertheless, the use of the risk assessment in the industrial environment (IMRA) is still an underdeveloped tool.

### **1.2 The exposure assessment as part of risk assessment**

The exposure assessment as a part of the risk assessment process estimates the presence of pathogens or microbial toxins and the probability that they will be present in the product at the time of consumption or at a given moment of the production process (process level for example), which is known as "farm to table" framework. The exposure assessment takes into account some factors such as the frequency of contamination caused by the pathogen or its level in the food during the shelf-life. At the same time, these factors can be influenced by:

- characteristics of the pathogen and its environment.
- microbial ecology of food and the initial contamination of raw material.
- level of hygiene control, methods of processing and preservation, packaging, distribution and storage.
- preparation of food: cooking, holding times, etc.
- consumption patterns.

All this make the levels of microbial pathogens in foods very dynamic. Therefore, the exposure assessment should describe the microorganism behaviour from production to consumption. In this sense, it is necessary to have instruments to facilitate the completion of the quantitative exposure assessment. Useful tools to develop a proper quantitative exposure assessment are predictive microbiology and the Monte Carlo simulation.

### **1.3 Predictive microbiology and mathematical models in exposure assessment**

Predictive microbiology is a discipline that combines elements of microbiology, mathematics and statistics to develop models that describe and predict the behaviour of microorganisms under certain experimental conditions. Specifically, it is based on the idea that organisms have reproducible behaviour and can be described as a function of different variables through a model. A model can be defined as: "a simplified representation, which includes key aspects of existing systems, which can be used for predictive purposes and control" (Eykhoff, 1979). Predictive microbiology is essential to perform a quantitative exposure assessment because through it and using a simulation procedure, changes in the number of microorganisms in a production line can be estimated providing an assessment of exposure

to a particular pathogen. Traditionally, inactivation and growth data have been adjusted to deterministic models obtaining in this way the kinetic parameters. However, the development of probabilistic models that actively consider product (ingredients or formulation) and process variability including the whole distribution (from minimum to maximum with all modes and percentiles) has gained importance in recent years and are essential in conducting risk assessment studies (Baert et al. 2009; Buchanan & Whiting, 1998; Hoornstra et al. 2001; Nauta, 2002).

#### **1.4 Simulation: an essential tool in exposure assessment.**

The simulation has been an important tool in industrial design tasks regardless of the application field. Although it has been considered the last alternative in the absence of data, recent advances in software have made it one of the most used and accepted tools in analysis and research. Simulation using computer allows us to replicate experiments or recreate scenarios using selective changes in the parameters or operating conditions. Correlations between sequences of random numbers to improve the statistical analysis of simulation output can be also entered. However, the simulation is an imprecise technique that provides only statistical estimates, not exact results, likewise, only compare alternatives rather than to generate an optimal value.

Because the sample of a particular distribution involves the use of random numbers, stochastic simulation is sometimes called **Monte Carlo simulation**. This method is one of the most powerful and used to analyse complex problems which may occur on sceneries related with food safety and is particularly appropriate in systems with many degrees of freedom (Larcher, 2006). It presents the advantage of being broadly applicable and relatively easy to use. It must be remembered that the Monte Carlo simulation is a computer-based technique that allows that variations to the input variables are propagated through inactivation, growth or germination mathematical models, which provides information about variation in the final result. In addition, Monte Carlo simulation allows also the possibility of carrying out a sensitivity analysis of all parameters involved in the model. The use of the Monte Carlo simulation, where input parameters are described as frequency distributions, is an example of stochastic or probabilistic analysis (van Gerwen and Gorris, 2004). However, the application of the Monte Carlo simulation at a process level exposure assessment study is still scarce and only a few studies are available (den Aantrekker et al. 2003; Ferrer et al. 2006, 2007; Pina-Pérez et al. 2010; Sampedro et al. 2010).

The technique can be implemented in many ways, the easiest option is an “add-in” for Microsoft Excel being the most commonly used @Risk (Palisade Corporation, 2004) and Cristal Ball (Decissioneering, 2005). Besides, there is also the possibility of using the numerical analysis software Matlab (The Mathworks Inc, 2006).

In this chapter a description of the Monte Carlo technique applied to industrial exposure assessment is given through a simple illustrative example of modelling the germination of *Bacillus cereus* spores and the simulation of the number of spores that can germinate at a given time of a process, considering different scenarios.

## **2. Case study: Exposure assessment of *Bacillus cereus* in liquid egg**

In the previous sections it have been discussed the possibility of using predictive microbiology and the Monte Carlo simulation for assessing the level of exposure of a hazard. In most cases, sporulated microorganisms or vegetative cells are used to conduct the

study. However, when sporulated microorganisms are used, germination has not been considered as a step and therefore there are only a few models describing this cellular event. The incorporation of information on germination as well as environmental factors that condition such a germination process, can be very valuable in estimating the risk associated with consumption of foods that can be contaminated with sporulated organisms.

In the example given in this chapter, the germination process of *Bacillus cereus* has been mathematically modelled and simulated in liquid whole egg. The Weibull distribution function was applied to model the germination process (Equation 1) by using a nonlinear regression performed with the SPSS statistic software (The Apache Software Foundation, 2000) and the Solver add-in option in Excel (Microsoft Corporation, 2003). In the last decade, the Weibull model has been often applied in food technology mainly in those cases where the inactivation of microorganisms did not follow the log-linear model. It has been used to describe degradation kinetics of food, those events can be considered failures in the system after being subjected to some stressful conditions during a certain time (Garcia, 2004). It has also been used to describe the kinetics of hydration of cereal used for breakfast (Machado et al., 1998) and thin-layer drying of cereals, grains and fruits (García, 2004). In the area of microbiology it has been applied to describe the survival kinetics of different microorganisms (Fernandez et al., 2002; Ruiz et al., 2002) to follow the combined effects of pH and temperature on the thermal resistance of *B. cereus* (Collado et al., 2003) and to compare the thermal resistance of *B. subtilis* (Jagannath et al., 2005). In the case of nonthermal technologies, it has been used to describe the inactivation kinetics of *E. coli* by pulsed electric fields processing, or by high pressure processing (Rodrigo et al., 2003), *L. plantarum* (Sampedro et al., 2006) and *C. sakazakii* (Pina-Perez et al. 2007a, Pina-Pérez et al., 2007b), among other applications.

$$\frac{G_t - G_0}{G_\infty - G_0} = 1 - e^{-\left(\frac{t}{\alpha}\right)^\beta} \quad (1)$$

where  $G_t$  is the germination at time  $t$  of the experiment,  $G_0$  is the initial germination,  $G_\infty$  is the germination at the equilibrium point,  $\alpha$  is the scale parameter, and  $\beta$  is the shape parameter of the Weibull distribution respectively.

Considering the results of modelling the germination, a simulation that considered the effect of induced germination and subsequent pasteurisation on the final population of *Bacillus cereus* in liquid egg was carried out by using equation 2. For that purpose, the Monte Carlo simulation was applied using the numerical analysis program Matlab 7 (The Mathworks Inc, 2006). It was considered that the time and the parameters  $\beta$  and  $\alpha$  of the Weibull distribution followed a uniform distribution.

$$\text{Log}N = \text{Log}N_g - \frac{t}{D_R * 10^{\left(\frac{T_p - T}{z}\right)}} \quad (2)$$

where  $N$  is the final population of microorganisms,  $N_g$  indicates the germinated population of *B. cereus*,  $t$  is time,  $D_R$  is the D value  $t$  65 ° C,  $T_p$  is the temperature of pasteurisation,  $T$  is the treatment temperature and  $z$  is the sensitivity of microorganism to the variation of temperature. In the simulation model, the variables  $t$ ,  $D_R$ ,  $T_p$ ,  $z$ , are not considered deterministic but probabilistic. Therefore, they do not enter the model as exact values but as probability distributions, which for simplicity are considered as uniform.

## 2.1 Modelling the germination of *Bacillus cereus*

Figure 1 shows the response of *Bacillus cereus* in liquid egg, after applying three different concentrations of inosine as germinant (10, 5 and 1 mM).

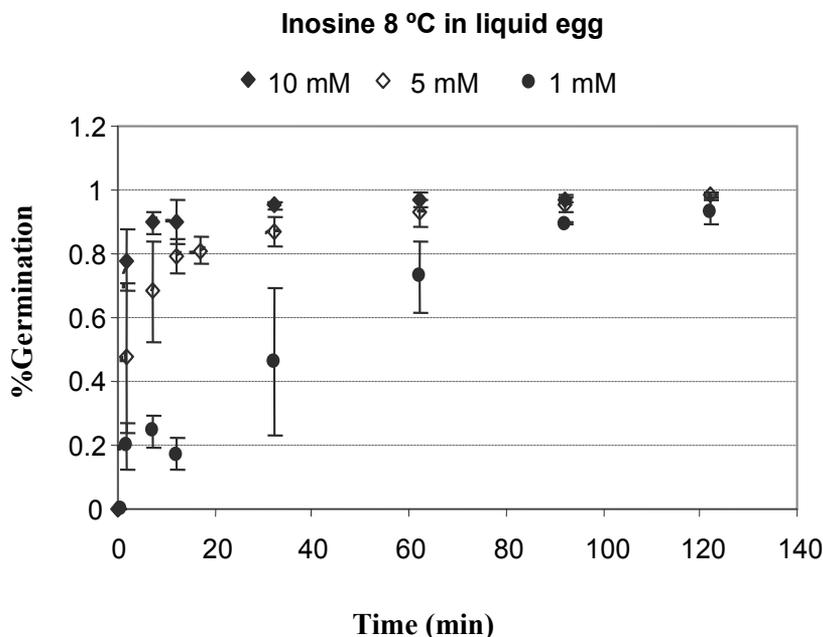


Fig. 1. Evolution of the germination of *B. cereus* at 8 °C in liquid egg using inosine as germinant.

It was found that at the temperature of 8 °C the endospores germinated well and after 122 minutes, germination was around 93 and 98%. On the other hand, a clear effect of the concentration of inosine on the germination response was observed. As increasing the concentration of germinant, germination was faster, existing significant differences between the three concentrations used in the study (10, 5 and 1 mM) according to Tukey's test (Sig <0.005).

From literature studies (Fernández et al. 2001) it can be concluded that endospores of *B. cereus* in liquid egg can survive the pasteurisation process thus can cause serious health problems. Since germination is considerable even at refrigeration temperatures, it could be convenient to apply this germination treatment previous to thermal pasteurisation. This technological procedure can help in reducing the number of endospores present in liquid egg. Therefore, the risk of illness caused by *B. cereus* could be reduced.

Figure 2 shows the experimental data fitted to the Weibull model according to equation 1.

As a measure of goodness of fit, the mean square error (MSE), the adjusted regression coefficient ( $R^2$ ) and accuracy factor ( $A_f$ ) were used. The MSE reports the error variance, so that the lower the value the better the model fits the data. With respect to  $R^2$ , the higher the value the better the model describes the data. Finally, the accuracy factor  $A_f$  provides a measure of the average difference between the values obtained experimentally and the values predicted by the model. The closer the value to the unit the better is the description of experimental data made by the model. Table 1 shows the values of MSE,  $R^2$  and  $A_f$  obtained for each non-linear regression.

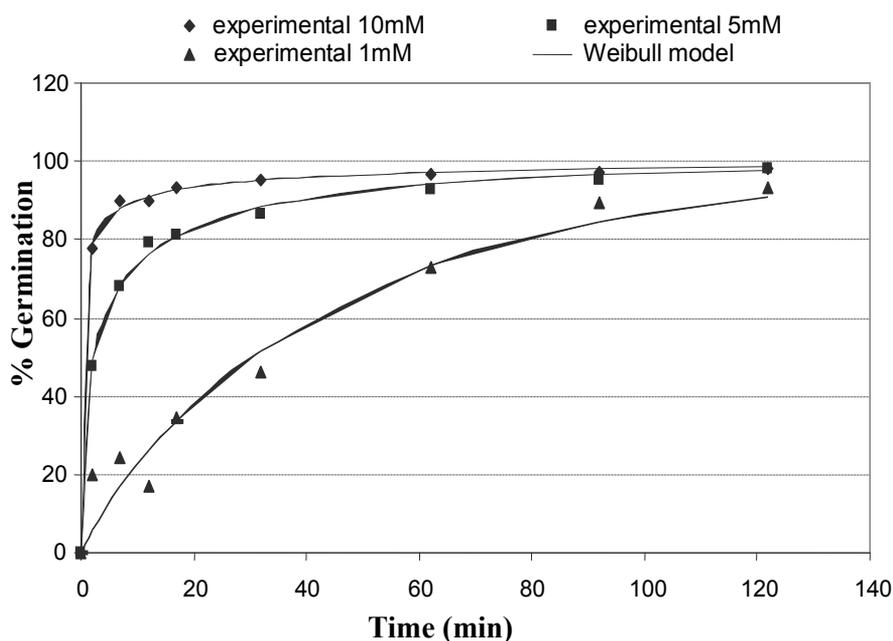


Fig. 2. Weibull fitted model to the germination data of *B. cereus* in liquid egg at 8°C.

Germinant concentration	$\alpha$	$\beta$	MSE	R <sup>2</sup>	A <sub>f</sub>
10 mM	0,331±0,243	0,243±0,136	0,890	0,998	1,007
5 mM	7,521±0,948	0,570±0,072	2,779	0,996	1,013
1mM	45,802±7,194	0,920±0,164	55,021	0,943	1,152

Table 1. Values and confidence intervals of the different parameters after fitting experimental data to the Weibull distribution function.

The results in Table 1 indicated that the model was adequate, especially for cases where the concentration of inosine was high. R<sup>2</sup> values obtained were close to 1, corresponding to their highest values with the lowest MSE and A<sub>f</sub>. A low value of MSE means a small error of variance, and therefore, the model fitted well the experimental data. The value of A<sub>f</sub> indicated that, on average, the prediction made by the model differs from the experimental data from 0.7 to 15% for 10 to 1 mM respectively. On the other hand, the value of *a* could be considered as the kinetic constant, which could be related to the rate at which germination occurs; the higher value of *a*, the slower the germination response. In fact, it can be seen that *a* increased with decreasing the concentration of germinant. Therefore, a high value of *a* would indicate a lower germination rate.

The  $\beta$  parameter describes the shape of the curve. Thus,  $\beta$  values greater than 1 ( $\beta > 1$ ) indicate the presence of shoulders, while  $\beta$  values lower than 1 ( $\beta < 1$ ) notes the appearance of tails. When  $\beta$  is equal to 1 ( $\beta = 1$ ), the Weibull distribution reduces to a first-order kinetic model. In this case study, all  $\beta$  values were lower than 1. As far as germination was concerned,  $\beta$  could be related with the saturation of the receptor to the inosine germinant. In

this context, the highest germination was achieved for smaller  $\beta$  values, reaching at the end of the study an germination equilibrium point. Contrarily, the curves shown higher  $\beta$  values corresponded to cases where final germination obtained was lower.

It can be concluded that the Weibull model fitted satisfactorily the experimental data, rendering predictive curves from which predictions could be obtained for a determined amount of treatment time. However, despite its usefulness, this is not sufficient to carry out an exposure assessment.

## 2.2 Monte Carlo simulation

The industrial exposure assessment requires a model that relates the probability distribution of the population at the end of the production process with the probability distribution of the population present at the beginning of the process. For this reason, a Monte Carlo simulation for each germinant concentration was conducted.

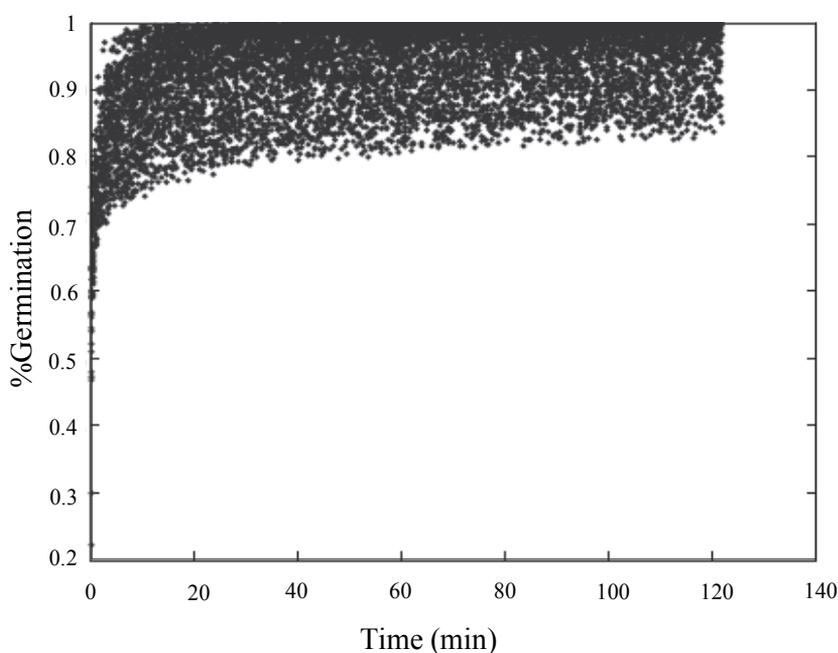


Fig. 3. Simulation of germination of *B. cereus* with 10 mM inosine at 8 °C in liquid egg.

The first variable considered by the simulation model was the initial concentration of microorganisms. This concentration was considered  $10^6$  cfu/g based on a study provided by producers in the different stages of an industrial production process. Specifically, it was studied the steps that could influence the presence of *B.cereus* in the final product, concluding that there was a risk of microbial contamination in the following operations: egg reception, cracking, separation and storage. In the case of the reception of eggs, the risks lied on the bacteria present both outside and inside the egg shell. Additionally, during the breaking and cracking, separation and storage, microbial concentration could increase in the liquid egg due to the contact with dirty egg-shells or contaminated surfaces and equipments.

Probability distributions of other variables to be considered in the germination simulation were introduced in the exposure assessment model. It was considered that the time, the

shape and scale parameters ( $\alpha$  and  $\beta$ ) and the germination rate followed a uniform distribution. That is, each value was associated with the same probability of occurrence in a known interval. In the case of the parameters  $a$  and  $\beta$ , that interval was calculated from the values provided by the model fit and their confidence interval.

The results after a simulation with 10,000 iterations are shown below. The curves obtained by simulation were consistent with those experimentally obtained. Figure 3 shows the result of germination simulation in the case of adding 10 mM of inosine. The possible germinative scenarios are contained in the cloud of points showed in the figure. The results for the other cases studied (1.5 and 1 mM), are presented in Figures 4 and 5.

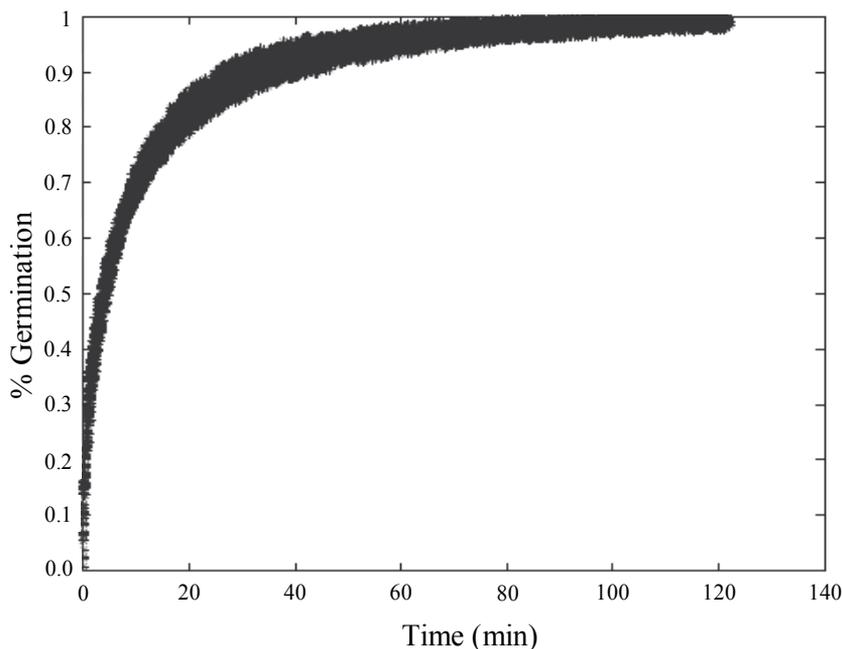


Fig. 4. Simulation of germination of *B. cereus* with 5 mM inosine at 8 °C in liquid egg.

The introduction of these results within a thermal treatment model is useful because it allows calculate the probability of occurrence of *B.cereus* between certain limits in the final product. Thus, it was calculated the probability that the liquid egg contained a number of microorganisms after the germination and the subsequent pasteurisation of samples (Table 2).

N° ufc/g	Probability Inosine 10mM	Probability Inosine 5mM	Probability Inosine 1mM
Between 0-1	100%	100%	100%
Between 10 <sup>5</sup> -5x10 <sup>5</sup>	0	0	0

Table 2. Probability that final product being contaminated with a certain number of microorganisms after receiving a standard pasteurisation process (10<sup>5</sup> is considered the infective dose).

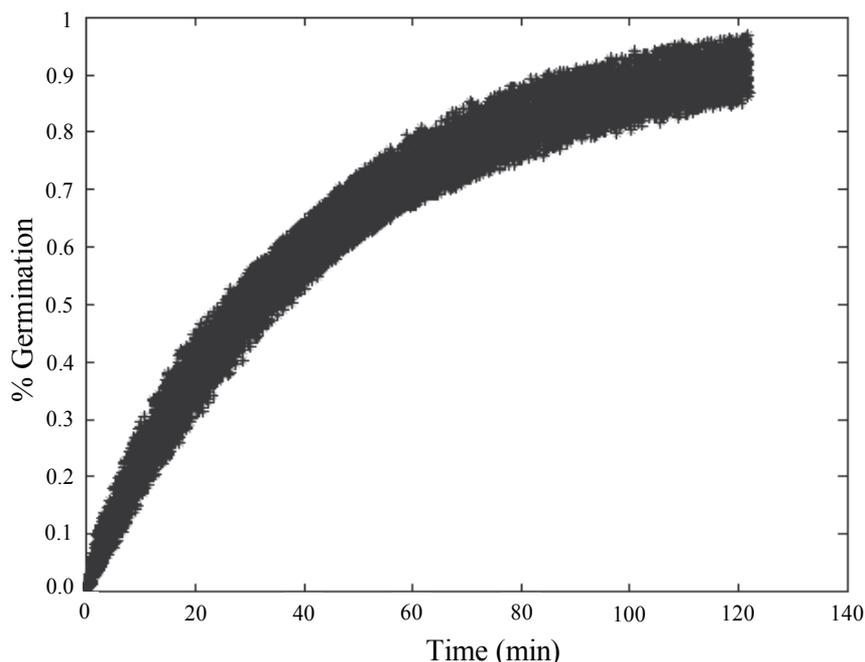


Fig. 5. Simulation of germination of *B. cereus* with 1 mM inosine at 8 °C in liquid egg

These results indicate that induced germination followed by a pasteurisation would yield a final product ready for storage and packaging, virtually *B.cereus* free. This is important for two reasons. First, it ensures product safety in the absence of cold chain due to a breakdown during storage, packaging, distribution and consumption. Secondly, the probability of rejection of contaminated liquid egg batches is considerably reduced because the maximum level allowed by manufacturing companies, cream caramel producers, is not attained. It should be noticed that most of the liquid egg produced is employed as raw material for manufacturing of desserts, e.g. cream caramel, with their own very restrictive internal quality standards, requiring absence of *B. cereus* in 400 g of liquid egg.

It is generally accepted that the infective dose by *B. cereus* is  $10^5$  cfu/g (Doyle, 2004). In this case, the probability of liquid egg could be contaminated at that level is practically zero, which would be within safety limits for its commercialisation.

What would happen with the final population of *B. cereus* in liquid egg if the scenario changes, for example, the occurrence of a failure in the pasteurisation process? Let us see the result considering the following scenario: the liquid egg receives a thermal treatment at 57°C, rather than the mandatory at 65°C or higher. Table 3 shows the results obtained.

N° Ufc/g	Probability Inosine 10mM	Probability Inosine 5mM	Probability Inosine 1mM
Between 0-1	0	0	0
Between $10^5$ - $5 \times 10^5$	15.40%	32.51%	51.80%

Table 3. Probability of final product contamination after a failure in the pasteurisation process ( $10^5$  is considered the infective dose).

After comparing the results of Table 3 with those presented in Table 2 we can conclude that when a failure in the pasteurisation process was produced, and the liquid egg was treated at 57 °C, the probability of finding *B.cereus* in the final product, in a concentration that can produce illness, would be considerably higher.

In resume, the result of the simulation shows that as the amount of germinant is reduced the probability of finding microbial counts within the range  $10^5$ - $5 \times 10^5$  (infective dose) increases. The result is consistent with the kinetic results since the 1 mM concentration produced less germination (Figure 1) and consequently more endospores will survive the thermal treatment at 57 °C. It would be advisable; therefore, the use of 10 mM as optimal concentration of germinant together with good hygienic practices during the production to prevent for a large number of *B. cereus* spores in the final product even if a failure in the pasteurisation process occurs. This simulation result warns about the need of a proper maintenance service of the pasteurisers and indicates that pasteurisation is a Critical Control Point in the manufacture of liquid egg.

It is important to note that the above findings are influenced by the boundary conditions applied. In fact, the simulation was performed by considering a very high initial contamination:  $10^6$  cfu/g (the worst scenario). In this sense, the simulation also demonstrates that the implementation of HACCP is a crucial activity to maintain the safety of the product, even when a failure in the process is produced. For that, the introduction of the Monte Carlo simulation by using an Industrial Risk Assessment approach is of great interest for the industry.

### 3. Conclusions

The results obtained with this simulation by using Monte Carlo are very useful since they give a quantitative estimate of the exposure assessment at the end of the pasteurisation process. This information can be crucial in establishing the management options for other operations that follow the pasteurisation process until the product reaches the consumer. It is important to point out that the simulation of the germination can be implemented as part of a process that simulates the entire production chain through the use of modules for each one of the stages, approach known as the Modular Process Risk Model (MPRM).

Achieving a MPRM is a complex task due to the many variables that may affect the evolution of the microbial counts. Basically, the MPRM is characterised by identifying the basic operations (modules) that comprises the food production and the conditions that determine the presence of the microorganism in those modules, and later on use mathematical models that explain adequately the situation under study. Finally, additional information such as the dose-response relationship, consumption patterns and handling of food in the household or the establishment of a given population is needed since the level of exposure depends not only on what happens during the process but also on the consumer behaviour. Taking into account the previous considerations, the usefulness of the simulation carried out in this case study is clear, since it makes possible its inclusion in a risk model with a modular structure, providing the basis for a quantitative industrial risk assessment.

This case study that simulates the germination of *B. cereus* for the subsequent inclusion in a heat treatment model, has led to different values of probability of having a certain number of microorganisms at the end of the pasteurisation process, considering different scenarios, assessing the exposure level to *B. cereus* in liquid egg. It has showed how the Monte Carlo Tool simulation is a tool to be considered in the food safety and decision making in an industrial plant, allowing management activities considering a quantitative approach.

#### 4. References

- Baert, K.; Francois, K.; De Meulenaer, B. & Devlieghere, F. (2009). Risk assessment: A quantitative approach. In: *Predictive modeling and risk assessment*, Costa, R., Kristbergsson, K. (Eds.), 21-22, Springer, 978-0-387-33512-4 New York:.
- Buchanan, R.L. & Whiting, R. C. (1998). Risk assessment: A means for linking HACCP plans and public health. *Journal of Food Protection*, 61, 1531-1534, 0362-028X
- Collado, J.; Fernández, A.; Cunha, L.M.; Ocio, M.J. & Martínez, A. (2003). Improved model based on the Weibull distribution to describe the combined effect of the pH and temperature on the heat resistance of *Bacillus cereus* in carrot juice. *Journal of food protection*, 66, 978-984, 0362-028X.
- Den Aantrekker, E.D.; Beumer, R.R.; van Gerwen, S.J.C.; Zwietering, M.C.; van Schothorst, M. & Boom, R.M. (2003). Estimating the probability of recontamination via the air using Monte Carlo simulations. *International Journal of Food Microbiology*, 87, 1-15, 0168-1605.
- Doyle, M.P. (2004). Bacteria Associated with Foodborne Diseases. *Institute of Food Technologists*. August Scientific Status Summary, 18-19.
- Eykhoff, P. (1979). *System identification: parameter and state estimation*. Ed. John Wiley and sons, 0-471-24980-7, pp: 555. New York.
- Fernández, A.; Ocio, M.J.; Fernández P.S. & Martínez, A. (2001). Effect of heat activation and inactivation conditions on germination and thermal resistance parameters of *Bacillus cereus* spores. *International Journal of Food Microbiology*, 63, 257-264, 0168-1605.
- Fernández, A.; Collado, J.; Cunha, L.M.; Ocio, M.J. & Martínez, A. (2002). Empirical model building based on Weibull distribution to describe the joint effect of pH and temperature on the thermal resistance of *Bacillus cereus* in vegetable substrate. *International Journal of Food Microbiology*, 77, 147-153, 0168-1605.
- Ferrer, C.; Rodrigo, D.; Pina, M.C.; Klein, G.; Rodrigo, M. & Martínez, A. (2007). The Monte Carlo simulation is used to establish the most influential parameters on the final load of pulsed electric fields *E. coli* cells. *Food Control*, 18, 934-938, 0956-7135.
- Ferrer, C.; Tejedor, W.; Klein, G.; Rodrigo, D.; Rodrigo, M. & Martínez, A. (2006). Monte Carlo simulation to establish the effect of pH, temperature and heating time on the final load of *Bacillus stearothermophilus* spores. *European Food Research and Technology*, 224, 153-157, 1438-2377.
- García Pascual, P. (2004). Deshidratación y conservación de setas silvestres: *Morchella esculenta* y *Boletus edulis*. PhD Dissertation. Universidad Politécnica de Valencia (Spain).
- Hoornstra, E.; Northolt, M.D.; Notermans, S. & Barendsz, A.W. (2001). The use of quantitative risk assessment in HACCP. *Food Control*, 12, 229-234, 0956-7135.
- Jagannath, A.; Tsuchido, T. & Membre, J.M. (2005). Comparison of the thermal inactivation of *Bacillus subtilis* spores in foods using the modified Weibull and Bigelow equations. *Food Microbiology*, 22, 233-239, 0740-0020.
- Larcher, L. & Cattaneo, C. (2006). Simulación de crecimiento de microorganismos utilizando el método de Monte Carlo. *Mecánica Computacional*, XXV, 2505-2518, 1666-6070.
- Machado, M.; Oliveira, F.; Gekas, V. & Singh, P. (1998). Kinetics of moisture uptake and soluble solids loss by puffed breakfast cereals immersed in water. *International Journal of Food Science and Technology*, 33: 225-237, 0022-1155.

- Nauta, M.J. (2002). Modelling bacterial growth in quantitative microbiological risk assessment: is it possible? *International Journal of Food Microbiology*, 73, 297- 304, 0168-1605.
- Pina-Pérez, M.C.; Rodrigo Aliaga, D.; Ferrer Bernat, C.; Rodrigo Enguádanos, M. & Martínez López, A. (2007a). Inactivation of *Enterobacter sakazakii* by pulsed electric field in buffered peptone water and infant formula milk. *International Dairy Journal*, 17, 1441-1449, 0958-6946.
- Pina-Pérez, M.C.; Rodrigo Aliaga, D.; Saucedo Reyes, D. & Martínez López, A. (2007b). Pressure inactivation kinetics of *Enterobacter sakazakii* in infant formula milk. *Journal of Food Protection*, 70, 2281-2289, 0362-028X.
- Pina-Pérez, M.C.; García-Fernández, M.M.; Rodrigo, D. & Martínez-López, A. (2010). Monte Carlo simulation as a method to determine the critical factors affecting two strains of *Escherichia coli* inactivation kinetics by high hydrostatic pressure. *Foodborne Pathogens and Disease*, 7, 459-466, 1535-3141.
- Rodrigo, D.; Barbosa-Canovas, G.V.; Martínez, A. & Rodrigo, M. (2003). Weibull distribution function based on an empirical mathematical model for inactivation of *Escherichia coli* by pulsed electric fields. *Journal of Food Protection*, 66, 1007-1012, 0362-028X.
- Ruiz, P.; Ocio, M.J.; Cardona, F.; Fernández, A.; Rodrigo, M. & Martínez, A. (2002). Nature of the inactivation of *Bacillus pumillus* spores heated using non-isothermal and isothermal treatments. *Journal of Food Science*, 67, 776-776, 0022-1147.
- Sampedro, F.; Rivas, A.; Rodrigo, D.; Martínez, A. & Rodrigo M. (2006). Effect of temperature and substrate on pulsed electric field (PEF) inactivation of *Lactobacillus plantarum* in an orange juice-milk beverage. *European Food Research and Technology*, 223, 30-34, 1438-2377.
- Sampedro, F.; Rodrigo, D. & Martínez, A. (2010). Modelling the effect of pH and pectin concentration on the PEF inactivation of *Salmonella enterica* serovar *Typhimurium* by using the Monte Carlo simulation. *Food Control*. (In press): <http://dx.doi.org/10.1016/j.foodcont.2010.09.013>, 0956-7135.
- Van Gerwen, S.J.C. & Gorris, L.G.M. (2004). Application of microbiological risk assessment in the food industry via tiered approach. *Journal of Food Protection*, 67, 2033-2040, 0362-028X.
- Zwietering, M.H. & Nauta, M.J. (2007). Predictive models in microbiological risk assessment. In: *Modelling microorganisms in food*, Brul, S.; Van Gerwen, S. & Zwietering, M.H. (Eds.), 110, CRC Press, 9780849391491. Boca Ratón.

# Monte Carlo Simulation of Radiative Transfer in Atmospheric Environments for Problems Arising from Remote Sensing Measurements

Margherita Premuda

*Institute of Atmospheric Sciences and Climate, National Research Council (ISAC-CNR),  
via Gobetti 101, 40129 Bologna,  
Italy*

## 1. Introduction

The way in which solar radiation distributes itself in the atmosphere and on the ground is well known. It is beyond the scope of this book and the reader can refer to more specific references (Kondratyev, 1969; Goody & Yung, 1995; Liou, 1998) for more detail. Solar radiation, essentially in the visible-ultraviolet frequency range, and infrared radiation, emitted by the terrestrial surface, are the prevailing energy sources for general atmospheric circulation. They are thus particularly important for meteorological and climatic studies. It would therefore be of great interest, for instance, to be able to calculate the influence of the presence of ozone and trace gases, water vapour and clouds, and various aerosols on radiative transfer and global thermal energy in the atmosphere or in particular regions of it.

These considerations naturally lead to the analysis of radiative transfer in the terrestrial atmosphere. This can be done using an atmospheric radiative transfer model (RTM) which also includes the possibility of single and multiple scattering events. Numerical and analytical methods can be used to solve a radiative transfer equation (Stamnes et al., 1988; Lenoble, 1977; Fouquart et al., 1980). A Monte Carlo approach is particularly suitable when multiple scattering significantly affects the results or where marked anisotropy of scattering and complex geometrical configurations are involved. The interest in such problems has increased through recently developed techniques related to remote sensing observations (satellite-based, ground-based or airborne) of the Earth's surface and atmosphere, involving the use of spectral radiation dispersion systems (mainly radiometers, spectrometers and interferometers) and active systems (principally RADAR, LIDAR and SODAR)

Among these various techniques, DOAS (Differential Optical Absorption Spectroscopy) and LIDAR (Light Detection And Ranging) investigations on the presence of particular atmospheric constituents or of atmospheric phenomena such as clouds, fog, rain, etc., are of special interest. With reference to surface remote sensing observations, for instance, the effects of atmospheric absorption and scattering constitute a noise element, which has to be evaluated by calculations.

Simulation of both LIDAR and DOAS systems deals with radiation in the UV/visible spectral range. For simulation purposes, a LIDAR system can be schematized as a pulsed

laser “disk” source of monochromatic radiation and a “disk” receiver (the “disk” schematization will be explained later), respectively with small emitting and receiving angles, i.e., with a narrow field of view (FOV). By means of Monte Carlo simulations, it is possible to analyze the time and spatial distributions of the backscattering radiation due to various atmospheric components. The equation for a LIDAR backscattering signal as well as further considerations relating to Monte Carlo simulations has been published (Pace et al., 2003; Coletti & Fiocco, 1980).

DOAS (Noxon, 1975; Platt et al., 1979; Platt & Perner, 1980; Roscoe et al., 1999) is an established remote sensing technique which identifies and quantifies the trace gases in the atmosphere taking advantage of their absorption structures in the near UV and visible wavelengths of the solar spectrum (passive DOAS). For simulation purposes, DOAS systems can be represented by a disk detector of solar radiation, with a relatively small diameter and a narrow FOV.

For our purposes, the atmosphere can be considered to consist of two different kinds of components: molecular (gases) and non-molecular (aerosol), which interact with radiation in different ways. For both components, a knowledge of their interaction coefficients and of the angular distributions of the scattered radiation (phase functions) is required. Appropriate average values of interaction coefficients for suitable altitude subdivisions have to be computed by taking into account the height variation of their density. Regarding molecular components, both continuum and line absorption are considered. The former accounts for reciprocal interactions between molecules of the same or other species whereas line molecular absorption is connected to rotational and rotational-vibrational transition frequencies, characteristic of each kind of molecule. Using theoretical models it is possible to find analytical expressions for line absorption coefficients as functions of such frequencies (Clough, et al., 1981; Kneizys et al., 1983a; Kneizys et al., 1984; Kondratyev, 1969). For radiation scattering by molecular components, the well-known expression of scattering coefficients and phase functions arising from Rayleigh theory can be adopted (Kondratyev, 1969). For non-molecular atmospheric components, scattering coefficients and phase functions can be derived from the Mie theory (Kondratyev, 1969; Lenoble, 1977) for particles whose dimensions are comparable to the radiation wavelength. Besides the exact Mie phase functions, several approximate formulae are available, the most suitable being the Henyey-Greenstein approximation (Kondratyev, 1969; Lenoble, 1977), based on a knowledge of the asymmetry factor  $g$  (average cosine of the scattering angle). As can be seen in Figure 1 (Tomasi & Paccagnella, 1986), the Mie phase functions may show marked anisotropy.

In the simulation of radiation transport through the atmosphere the phenomenon of refraction has to be taken into account as a consequence of different refractive index values between contiguous geometrical shells. This phenomenon can be remarkable, for instance, in the case of vertical view detectors receiving solar radiation at solar zenith angles near to or greater than  $90^\circ$  (horizon) as can be seen in Figure 2, where a plot of a sun ray, obtained using the MOCRA (MOnTe Carlo Radiance Analysis) code for a  $93^\circ$  solar zenith angle (Premuda et al., 2009) is shown. This phenomenon is taken into account according to Snell's refraction law  $n_i/n_r = \sin \vartheta_r/\sin \vartheta_i$ , where  $n_i$  and  $n_r$  are the average refractive indices of the two contiguous regions involved and  $\vartheta_i$  and  $\vartheta_r$  are the incident and refracted angles with respect to the normal to the boundary surface, respectively. It should be noted that total reflection occurs for incident angles greater than a critical value.

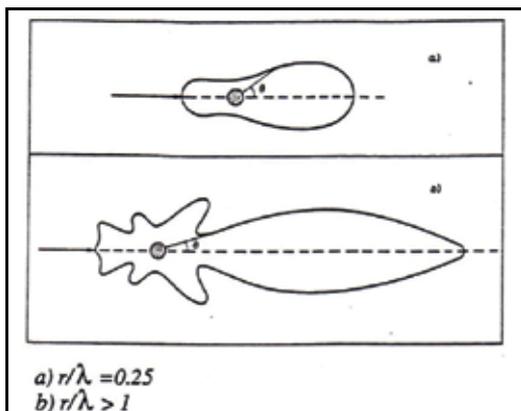


Fig. 1. Phase function for spherical particles. Reproduced by courtesy of SIF (Italian Physical Society) (Tomasi & Paccagnella, 1986).

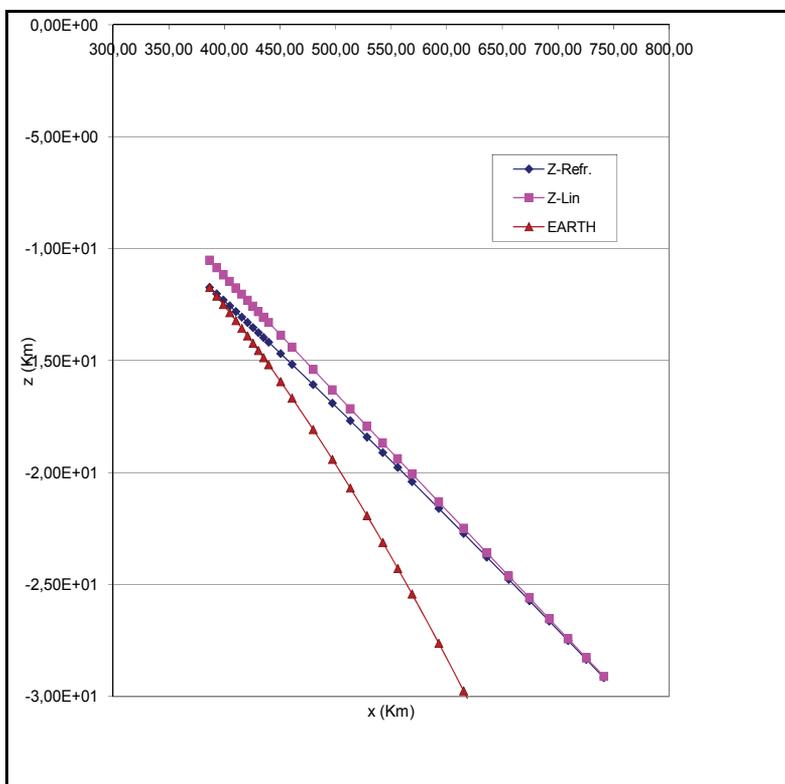


Fig. 2. Plot of refraction of a sun ray obtained using the MOCRA code for a  $93^\circ$  solar zenith angle. The refracted ray (dark blue) hits the Earth's surface (brown), whereas the linear ray, without refraction (magenta), would reach the vertical above the observer.

At ground level, the albedo phenomenon usually has to be taken into account according to appropriate surface albedo coefficients. Towards this end, the Lambert reflection law is frequently adopted, which assumes a cosine distribution law for the reflected radiation with

respect to the normal to the ground surface. The values of the albedo coefficient may significantly depend upon the radiation wavelength and on the surface characteristics (snow cover, foliage, bare soil, etc.). Bidirectional reflectance distribution functions (BRDF), particularly useful for analyzing ground reflectance properties, may also be considered, as they take into account the solar zenith angle, the observer line of sight and the azimuth between the solar and observer directions. A theoretical model of the BRDF compared to experimental observations is available (Walthall et al., 1985). A Monte Carlo simulation has also been used (Richtsmeier & Sundberg, 2009).

Very accurate molecular and non-molecular models are adopted in the radiance-transmittance MODTRAN codes (Berk et al., 1989; Kneizys et al., 1996; Acharya et al., 1998; Berk et al., 1999), where physical phenomena involving radiation in the infrared-ultraviolet spectral range are analyzed in detail. They are used widely by the remote-sensing community to model spectral absorption, transmission, emission and scattering characteristics of the atmosphere.

Through a knowledge of interaction coefficients and phase functions, it is possible to set up sequences of discrete and continuous probability distribution functions (p.d.f.) which allow Monte Carlo photon-history tracking.

Monte Carlo simulations related to LIDAR and DOAS systems will be described below, together with proper variance reducing techniques. A variety of Monte Carlo applications have been described (Marseguerra & Zio, 2002).

## 2. General features of Monte Carlo radiative transfer simulations in atmospheric environments

An appropriate description of a numerical or statistical simulation of radiation transport in an atmospheric environment requires a theoretical equation which governs the transport phenomenon to be stated and the physical and geometrical properties of the environment defined. To this end, in what follows, a simplified form of the integral transport equation will be given together with a possible environmental representation suitable for the simulation process, which will be subsequently described in its general fundamental lines.

### 2.1 Integral radiative transfer equation and Monte Carlo simulation

To analytically describe, in its photonic representation, radiation transport in atmospheric environments for time independent problems, the following integral form of the Radiative Transfer Equation (RTE) can be written for the specific intensity of radiation  $I(r, \Omega)$  defined as the photonic distribution function times  $ch\nu$ , where  $c$  is the velocity of light in a vacuum,  $h$  Planck's constant and  $\nu$  the radiation frequency

$$I(r, \Omega) = \Gamma(r_s, \Omega) \exp \left[ - \int_0^{|r-r_s|} ds'' k(r-s''\Omega) \right] + \int_0^{|r-r_s|} ds' Q(r-s'\Omega, \Omega) \exp \left[ - \int_0^{s'} ds'' k(r-s''\Omega) \right] \quad (1)$$

where  $k$  is the total extinction coefficient, sum of the scattering  $k_s$  and the absorption  $k_a$  coefficients.  $Q$  is the source density for emission and scattering defined as

$$Q(r, \Omega) = S(r) + \int_{4\pi} d\Omega' k_s(r, \Omega \cdot \Omega') I(r, \Omega'). \quad (2)$$

where  $I(r_s, \Omega)$  is the source radiation specific intensity. The  $S(r)$  term in equation (2) represents an internal radiation source.

As can be seen, the specific intensity of radiation in direction  $\Omega$  at point  $r$  is determined by the sum of two contributions: the first is the direct source contribution, due to the source specific intensity of radiation starting at point  $r_s$  and the second accounts for emission and scattering of the beam radiation coming from each element  $ds'$  of the path from  $r_s$  to  $r$  along the  $\Omega$  direction. Both contributions at point  $r$  are exponentially attenuated as a consequence of absorption and scattering collisions. It should be noted that in solar radiance analysis the first contribution accounts for the irradiance and the second for single and multiple scattering.

The integral form of the transport equation is suitably solved by means of numerical or Monte Carlo simulations (Marseguerra & Zio, 2002). It can be obtained directly from the more general integro-differential radiative transfer equation (Spiga et al., 1992; Premuda & Palestini, 1982; Premuda, 1994).

The attenuating exponentials are the same used to evaluate path lengths between collisions in Monte Carlo simulations. The length relevant to exponential attenuation of both contributions is the optical depth, defined as follows:

$$\tau(r, r_s) = \int_0^{|r-r_s|} ds'' k(r - s''\Omega). \quad (3)$$

In the standard “forward” Monte Carlo simulation, histories of photons emitted from the radiation source are followed until they hit the detector or disappear from the system as a consequence of leakage or absorption. Problems in this simulation procedure inevitably arise when small detectors with narrow FOVs are considered in the presence of an external broad source, due to the very low probability of a photon reaching the detector. To overcome this kind of problem, a “backward” Monte Carlo simulation is usually performed, taking into account that the linearity of the transport Boltzmann equation and the reciprocity relationship between the Green function  $G(P \rightarrow P_d)$  and its adjoint  $G^*(P_d \rightarrow P)$  make it possible to write for the flux  $\Phi(P_d)$ :

$$\Phi(P_d) = \int G(P \rightarrow P_d) S(P) dP = \int G^*(P_d \rightarrow P) S(P) dP \quad (4)$$

in which the source  $S(P)$  is always evaluated at the source point  $P$  (a detailed analysis of the Monte Carlo simulation of the adjoint transport Boltzmann equation is available (De Matteis & Simonini, 1978a; De Matteis & Simonini, 1978b)). In the case of solar sources, each photon history is therefore traced through the atmosphere as being generated by the detector according to its line of sight direction, taking into account the phase function for the angle which leads from a collision point  $P_c$  to the source point  $P$  on the external atmospheric boundary along the solar radiation direction. A backward simulation description for radiation transport in the atmosphere, where polarization effects are also taken into account, is available (Collins et al., 1972). In atmospheric radiative transfer simulations, the backward scheme is even more suitable, because in photon interactions with atmospheric constituents energy changes do not occur. A forward Monte Carlo radiative transfer simulation for photon tracing in three-dimensional cloudy atmospheres is foreseen in the Mystic code (Mayer & Kylling, 2000). Topography and an inhomogeneous surface albedo are considered.

## 2.2 The atmospheric environment and photon history tracking

For a radiation of assigned wavelength, the photonic interpretation allows the application to radiative transfer of simulation techniques usually adopted for particle transport in assigned materials. Starting from the source, trajectories of individual photons are followed, according to the physical and geometrical properties of the environment of interest and taking into account the discrete and continuous probability distribution functions belonging to possible events for the photon.

A synthetic description of the essential characteristics of the atmospheric environment required for the simulation process and of the standard "forward" history tracking is given below.

### *a) The atmospheric environment*

Regarding the physical properties, as mentioned previously, reaction coefficients and phase functions for both molecular and aerosol particles are required. A knowledge of these parameters, together with refractive indices, which characterize each kind of atmosphere, can be profitably used to build a library data set, then used as the source of the fundamental physical and geometrical values needed for the Monte Carlo simulation. In analyzing radiative transfer in atmospheric environments, it can be crucial to verify the effects on radiation transport of perturbations in some atmospheric constituents (e.g. small variations in ozone or carbon dioxide concentrations). In Monte Carlo simulations it is possible to evaluate these effects by considering simultaneously reference and perturbed environments. This can be done by tracing the photon histories in the reference environment, taking into account the perturbations using appropriate weighting functions. This simulation tool makes it possible to evaluate small effects which could be masked by statistical errors when using separate calculations. This is possible, for instance, in the MOCRA code where several perturbed environments can be considered simultaneously.

In radiation transport simulations in atmospheric environments, great advantage can be gained by proper representation of general 3D spherical multi-region geometries. The atmosphere can, for instance, be subdivided into cones, spherical shells and azimuthal half planes, obtaining very detailed geometrical descriptions of the various surface and atmospheric regions (Cupini et al., 2005). This allows one to take into account various ground altitudes and atmospheric profiles for each region, thus allowing for soil orography and latitudinal and longitudinal variations in the composition of the atmosphere. This is also possible with MOCRA and has been used to simulate the presence of an obstacle in horizontal passive DOAS measurements (Premuda et al., 2009).

When several kinds of atmosphere have to be simultaneously taken into account in the same calculation, as may occur, for instance, in general 3-D geometries, the data belonging to each of them must obviously be available in the library. The appropriate build-up and treatment of the physical and geometrical library data can make it possible, for instance, to foresee effects due to the injection of special aerosols (e.g., fumes) in a given atmospheric region with already assigned aerosols.

In the following Km are the units of length and, consequently,  $\text{Km}^{-1}$  are reaction coefficient units. As already pointed out, reaction coefficients, given by the product of the density and the proper particle reaction "cross section", will depend on the particle density behaviour along the vertical z-axis starting from the ground altitude, so that average values are to be computed in Monte Carlo simulations. To this end, the total assumed atmospheric height

range ( $z_0, z_{\max}$ ) can be subdivided into an assigned number NL of conveniently chosen geometrical layers. With reference to molecular particles, if one assumes an exponential behaviour of the density spatial distribution within each layer, depending upon the pressure and temperature profiles, one obtains, for the j-th layer:

$$\rho(z) = \rho(z_j) \exp\left[-(z - z_j) / H_j\right]; \quad H_j = (z_{j+1} - z_j) / \ln(\rho_j / \rho_{j+1}) \quad (5)$$

The average density value  $\langle \rho \rangle_j$  within the layer will be given by:

$$\langle \rho \rangle_j = \int_{z_j}^{z_{j+1}} \rho(z) dz / (z_{j+1} - z_j) \quad (6)$$

i.e.:

$$\langle \rho \rangle_j = (\rho_j - \rho_{j+1}) / \ln(\rho_j / \rho_{j+1}) \quad (7)$$

If, for numerical reasons, this relationship cannot be applied, the value  $\langle \rho \rangle_j$  can be assumed to be equal to the arithmetic mean of  $\rho_j$  and  $\rho_{j+1}$ . Similar relationships will hold, consequently, for the average molecular scattering  $\langle k_{Ms} \rangle_j$  and absorption  $\langle k_{Ma} \rangle_j$  coefficients. It can be observed that, although a unique value is usually given for the scattering coefficient, corresponding to that of air considered as a fictitious molecule with a molecular weight  $M_{\text{Air}} = 28.964$ , average absorption coefficients must be assigned for each molecular species. The average coefficient  $\langle k_{Ma} \rangle_j$  has to be interpreted as the sum over the average coefficients belonging to each of them.

Analogous considerations can be made for the average aerosol coefficients  $\langle k_{As} \rangle_j$  and  $\langle k_{Aa} \rangle_j$  for the same layers, and for the refraction indices. It can be seen that very different kinds of aerosols can be found along the z-axis, from the boundary layer to the high stratosphere, so that marked discontinuities can occur in the average reaction coefficients between contiguous layers.

For the j-th vertical layer, the following total coefficients can be considered:

$$\langle k_M \rangle_j = \langle k_{Ms} \rangle_j + \langle k_{Ma} \rangle_j; \quad \langle k_A \rangle_j = \langle k_{As} \rangle_j + \langle k_{Aa} \rangle_j. \quad (8)$$

According to eq. (3), the optical depth of the layer will be:  $\tau_j = \langle k \rangle_j h_j$ . This optical depth characterizes the layer in the atmospheric environment and gives a measure of the transparency of the layer to the incident radiation. For the sake of simplicity, in what follows the brackets will be omitted.

In the photon diffusion process, an essential role is played by molecular Rayleigh scattering, which affects the distance travelled and the motion direction following a collision. A theoretical formula for the scattering coefficient  $k_{Ms}$  which takes into account the dependence on wavelength  $\lambda$ , refraction index n and molecular number density N is given by Kondratyev (Kondratyev, 1969) together with phase function and refraction index expressions. A refraction index formula for standard air, which takes into account the wavenumber dependence, is available (Edlén, 1966). In the LOWTRAN-MODTRAN codes the following formula is adopted (Kneizys et al., 1983b):

$$(n-1) \cdot 10^6 = \left[ a_0 + a_1 / \left( 1 - (v/b_1)^2 \right) + a_2 / \left( 1 - (v/b_2)^2 \right) \right] * \frac{(P - P_w)}{P_0} * (T_0 + 15.0) / T + \quad (9)$$

$$- \left[ c_0 - (v/c_1)^2 \right] * P_w / P_0$$

where  $P_w$  is the water vapour pressure,  $P_n = 1013.25 \text{ mb}$ ,  $T_n = 275.15 \text{ K}$ ,  $a_0 = 83.43$ ,  $a_1 = 185.08$ ,  $a_2 = 4.11$ ,  $b_1 = 1.14 \cdot 10^5$ ,  $b_2 = 6.24 \cdot 10^4$ ,  $c_0 = 43.49$ ,  $c_1 = 1.70 \cdot 10^4$ .

A molecular scattering formula, which takes into account the depolarization coefficient  $\delta$ , is given by (Kneizys et al., 1984):

$$k_{Ms}(\lambda) = \frac{24\pi^3}{N\lambda^4} \left[ \frac{n^2 - 1}{n^2 + 2} \right]^2 \frac{6 + 3\delta}{6 - 7\delta} \quad (10)$$

It is stressed that the Rayleigh scattering coefficient rapidly increases as the radiation wavelength decreases. For the depolarization coefficient, a value of 0.0279 is assumed, for instance, in the MODTRAN code. In this code, the following formula is used for the phase function  $p(\theta)$  (Kneizys et al., 1983b):

$$p(\theta) = \frac{3}{16\pi} \frac{2}{(2 + \delta)} \left[ (1 + \delta) + (1 - \delta) \cos^2 \theta \right] \quad (11)$$

With reference to the aerosol reaction coefficients and phase functions, as mentioned above, the Mie theory can be applied, which holds for spherical homogeneous particles with radii  $r$  comparable to the wavelength  $\lambda$  of the incident radiation. The Mie theory is much more complex than Rayleigh theory, from both analytical and numerical points of view, and will not be described here. It will only be observed that, for a particle of radius  $r$ , the size parameter  $\rho = 2\pi r / \lambda$  and complex index of refraction  $m = n - iq$  are quantities of interest in the theory. If a distribution  $N(r)$  is available (see, for instance, Shettle, 1989 for possible distribution functions), average reaction coefficients over the distribution can be obtained. Regarding the phase function, approximate equations have been proposed. These include the well-known Henyey-Greenstein phase function, given by

$$p(\vartheta) = \frac{1}{4\pi} \frac{1 - g^2}{(1 + g^2 + 2g \cos \vartheta)^{3/2}} \quad (12)$$

where  $g$ , the asymmetry factor, is the average cosine of the scattering angle, i.e.:

$$g = \iint_{4\pi} p(\vartheta) \cos \vartheta d\Omega. \quad (13)$$

In MODTRAN, special models for different kinds of rain droplets are also considered.

The aerosol phase functions can be made available in the library by means of tables of values over assigned angles in the range ( $0^\circ - 180^\circ$ ). For the molecular phase functions, the formula given by equation (11) can be used directly during the simulation. Attention must be paid in sampling scattering angles: for molecular scattering, the scattering angle cosine can be sampled using equiprobable cosine tables; for aerosol scattering, on the other hand,

such tables are, as a rule, not possible to set up, due to the marked anisotropy in their phase functions (Premuda, 1994).

*b) Photon history tracking*

Once the physical and geometrical properties of the environments required for the simulation process have been established, the history tracking can be carried out as usual, taking into account the peculiarities of the problem to be solved. In this regard, it can be observed that in the efficiency of the simulation aimed at the evaluation of a given quantity a crucial role is played by the estimator chosen, which will assume a statistical value  $\varepsilon_i$  at the end of  $i$ -th history. If  $N$  histories have been processed, its average value  $\mu$  :

$$\mu = \frac{\sum_{i=1}^N \varepsilon_i}{N} \quad (14)$$

will give the estimate of the quantity of interest. As for the statistical error, this can be estimated by means of the standard deviation for both the estimator,  $\sigma_\varepsilon$  and the estimate,  $\sigma_\mu$ , through the well-known variance equations  $V_\varepsilon = \sigma_\varepsilon^2$  and  $V_\mu = \sigma_\mu^2$  :

$$V_\varepsilon = \frac{\sum_{i=1}^N (\varepsilon_i - \mu)^2}{N - 1}, \quad V_\mu = \frac{V_\varepsilon}{N} \quad (15)$$

A quantity used to evaluate the efficiency of the simulation process adopted is given by the so-called Figure of Merit (FOM):

$$FOM = \frac{1}{t\sigma_\mu^2} \quad (16)$$

where  $t$  is a measure of the computation time. As an example of the various possible choices of the estimator in the simulation process, one can consider the problem of finding the absorption by a given molecule, for instance ozone, of the radiation at an assigned wavelength in a layer. Two kinds of estimators can be adopted: the first one could be derived from the probability of the radiation being absorbed in a collision within the layer (collision estimator); the second by multiplying the distance travelled by the photon within the layer by the absorption coefficient of the molecule (distance estimator). This latter estimator can be conveniently applied when very few collisions occur within a frequently crossed layer.

To take into account in the simulation process the analytical information available on the physical events which may occur and to develop tools for enhancing the contribution of the photon history to the final result, a statistical "weight"  $w$  is associated with the travelling photon, taking an initial value  $w_0$  assigned at the beginning of the history. In the solar radiance analysis, for instance,  $w_0$  could be assumed to be equal to the radiation intensity  $I_0$  of the solar source. Regarding the photon source, the spatial and angular distributions of the emitted photons will make it possible to select for each history the starting parameters needed for the subsequent realization of the chain of possible events undergone by the photon in a given reference system. For this purpose, the importance must be stressed of the

availability of a proper generator of random numbers uniformly distributed over the interval  $(0, 1)$ , usually an arithmetic congruence, which, as much as possible, avoid internal correlations among the generated numbers (Knuth, 1981; De Matteis & Pagnutti, 1988). It is well known that discrete  $(p_1, p_2, p_3, \dots)$  probability distributions and continuous probability density functions (p.d.f.), both normalized to 1, have to be handled in the history tracking. In the first case, it is easy to select the kind of event from the distribution by means of proper comparisons of a chosen random number  $r$  with cumulative probability values  $(r < p_1; r < p_1 + p_2; r < p_1 + p_2 + p_3; \dots)$ . In the second, if  $f(x)$  is the p.d.f. normalized to 1 over the interval  $(a, b)$ , to obtain a value of  $x$  the following equation must be solved for the chosen random number  $r$ :

$$r = \int_a^x f(t) dt \quad (17)$$

From this equation, for instance, a value  $\tau$  of the optical path travelled by the photon is easily obtained from  $e^{-x}$  attenuating exponential function, normalized to 1 over the interval  $(0, \infty)$ . When a direct solution is not available or too expensive, special techniques (such as the so-called "rejection technique") can be used to obtain values of the variable  $x$  obeying the p.d.f. of interest. In any case, discrete cumulative probability distribution tables can be obtained by means of analytical or numerical integration for assigned values of the independent  $x$  variable:

$$P_i = \int_a^{x_i} f(t) dt \quad (i = 1, 2, \dots, N) \quad (18)$$

A value of  $x$  can be obtained from eq. (17), by means of proper linear interpolations for the chosen random number  $r$ . This is, for instance, the case of the Mie phase function, for which, as previously said, tables of values are assumed to be available for the diffusion angle or, more conveniently, its cosine. Taking into account, as already emphasized, the possible marked anisotropy of this distribution function, tabulating points are assumed to be accurately chosen to avoid improper choice of the scattered photon direction. Regarding the Rayleigh phase function, the same approach can be adopted. In this case, as already indicated, a given number (preferably a power of 2, e.g. 32) of cosine intervals can be derived, by means of interpolations, from the cumulative probability table, corresponding to equal probability intervals. In this way, a fast random access to the cosine table allows one to obtain the diffusion angle.

On the basis of the previous considerations, to trace, in its fundamental lines, a photon history, an initial assigned weight  $w = w_0$  and a selected starting point and motion direction must first be assigned, as stated previously. The optical path  $\tau = -\ln(r)$ , with  $r$  random number ( $r$  and  $1-r$  have the same distribution) to be travelled is then chosen and the corresponding distance  $d = \tau/k$  is computed, being  $k$  the reaction coefficient of the starting layer. If the distance  $d$  is exhausted within the layer, a collision takes place. A new weight for the photon is computed by means of the relation:  $w = w \frac{(k_{As} + k_{Ms})}{k}$ . According to the

probabilities  $p_{As} = \frac{k_{As}}{k_{As} + k_{Ms}}$  and  $p_{Ms} = \frac{k_{Ms}}{k_{As} + k_{Ms}}$  the choice between the scattering by an

aerosol or by a molecule is made. If a collision with an aerosol particle occurs and various kinds of aerosols exist within the layer, each with its proper phase function, with probabilities:  $p_{A1s} = \frac{k_{A1s}}{k_{As}}$ ,  $p_{A2s} = \frac{k_{A2s}}{k_{As}}$ , ..., the specific kind of aerosol involved will be

selected. On the basis of the proper diffusion cosine distribution function, a new motion direction will be chosen and a new path started. If the distance  $d$  to be travelled is greater than the distance from the starting point and the boundary of the layer along the motion direction and a leakage phenomenon from the geometrical system does not occur, the optical path exhausted within the layer will be computed, the remaining one being utilized for the next path in the new layer. If refraction is foreseen, according to the refraction indices  $n_1$  and  $n_2$  belonging to the two contiguous layers, the new refracted motion direction will be computed. It must be taken into account that a total reflection will occur if  $n_1 > n_2$  and, moreover, the angle of the incident direction on the boundary between the two layers is greater than the  $\theta_L$  angle, being  $\theta_L = \arcsin \frac{n_2}{n_1}$ . If the path to be traced hits the ground

surface and an albedo coefficient  $\alpha$  is foreseen, a photon with a weight  $w = \alpha * w$  will be reflected with a motion direction chosen according to the diffusion law  $f(\Omega) = \frac{2\cos\theta}{2\pi} = \frac{\cos\theta}{\pi}$

with respect to the normal direction to the ground surface (Lambert law).

For termination of the history tracking, various criteria can be adopted in addition to that from the leakage of the photon from the geometrical system. Among them, the following can be considered. 1) Assigning as input parameter a minimum weight fraction,  $w_{min}$ , the history will end if  $\frac{w}{w_0} < w_{min}$ . The minimum weight fraction has to be carefully chosen to

avoid expensive calculations giving non-essential contributions to the estimator. A value of  $10^{-5}$  could be, for instance, adopted in many cases. The sum over the histories of the unprocessed weights will give a measure of the bias introduced. 2) Having assigned a cutting value  $w_{cut}$ , if, after a collision,  $\frac{w}{w_0} < w_{cut}$ , the photon weight will not change. When

a subsequent collision occurs, with a probability  $p_s = \frac{k_{As} + k_{Ms}}{k}$  this decides whether scattering occurs. If this is not the case, the history tracking will end. A  $w_{cut}$  value equal to

0.2 could be, for instance, assigned. 3) Analogously to the previous point, if  $\frac{w}{w_0} < w_{cut}$  a

“russian roulette” game is played: with a probability  $p_r = \frac{w}{w_0}$  that the photon will survive

and the initial weight  $w = w_0$  will be restored; with a probability  $1 - p_r$  that the photon will be killed and the photon weight lost. At the end of all the histories processed, the sum of the total weight gained should be statistically equal to the total weight lost. To avoid large jumps in the weight, which can affect the variance,  $w_{cut} = 0.5$  could be a reasonable value to be assigned. In MOCRA, all three options are available.

According to the quantities to be estimated through the simulation process, analytical contributions can conveniently be computed during the history tracking, so as to reduce as

much as possible the variance of the calculation. Regarding, for instance, the molecular absorption in a layer, an example is given by the distance estimator previously described. A further estimate frequently required is given by the leakage from the geometrical system. At each collision point the optical distance  $\tau_e$  to the external boundary along the direction of motion is computed. The escape weight  $w_e = w e^{-\tau_e}$  will be the analytical contribution of the collision point to the searched-for estimate. The total sum of the analytical contributions will give the estimator value for that history.

In addition to analytical evaluations during the history tracking, variance reducing techniques can be devised which alter the natural sequence of the events undergone by the photon, locally or over the entire physical and geometrical system, giving rise to more efficient unbiased contributions to the required estimates. The most common is given by the well-known forced collision technique. If the direction of motion from a starting or a collision point does not cross a reflecting boundary, the leakage optical distance  $\tau_e$  is computed as previously described. The two possibilities for the photon escaping from the system with a probability  $e^{-\tau_e}$  or colliding within it with a probability  $1 - e^{-\tau_e}$ , are taken into account separately. More precisely, a photon with a weight  $w_e = w e^{-\tau_e}$  is assumed to travel within the system and escaping from it, whereas a photon with a weight  $w_c = w(1 - e^{-\tau_e})$  is assumed to collide. It should be noted that both photons are to be considered when a distance estimator for the absorption is envisaged. To choose the

collision point, the p.d.f.  $f(x) = \frac{e^{-x}}{1 - e^{-\tau_e}}$ , normalized to 1 over the interval  $(0, \tau_e)$ , will then

be adopted for a photon of weight  $w = w_c$ . When this technique is applied, the first criterion among those previously described to end the history tracking has to be used. In this case, as a precaution, a maximum number of collisions per history can be assigned. It must be observed that the forced collision technique is an expensive one and it may cause, moreover, considerable changes in the travelling photon weight which can affect the variance. It should be used when the collisions within a system of small optical thickness are of special interest. In any case, the FOM should be taken into account.

Further special variance-reducing techniques will be described within the context of the following LIDAR and DOAS simulation analysis.

As previously highlighted, to analyze the effect of environmental perturbations on the final result due to small changes in physical parameters which do not affect the phase functions, a simulation process which takes into account, simultaneously, the unperturbed (reference) environment and the perturbed one can be adopted. As an example of application, the Air Mass Factor calculation (described in section 3.2) according to its theoretical definition can be considered, where the solar radiances with and without the absorption of a particular molecule are required.

In its basic lines, the procedure can be derived from the so-called "importance sampling" technique: given an assigned p.d.f.  $f(x)$ , normalized to 1 over  $(a, b)$  to calculate the average value of a function  $h(x)$  over the distribution  $f(x)$ , i.e.:  $\langle h(x) \rangle = \int_a^b h(x)f(x)dx$ , a more suitable p.d.f.  $g(x)$  can be adopted, by using the weighting factor  $\frac{f(x)}{g(x)}$  to take into

account the change in the distribution. If  $h'(x) = \frac{f(x)}{g(x)}h(x)$  then:  $\langle h(x) \rangle = \int_a^b h'(x)g(x)dx$ .

According to this device, a photon weight will be associated with the perturbed environment. The tracking of the photon history will be carried out in the reference system, taking into account the perturbed one by updating its weight according to proper weighting factors. As an example, if, in the reference system, along its line of flight a photon crosses a boundary, a weighting factor  $w_f = e^{-\delta\tau}$ ,  $\delta\tau = \tau' - \tau$ , has to be used, where  $\tau$  and  $\tau'$  are the optical paths from the starting point to the boundary for the reference environment and for the perturbed one, respectively. If a collision occurs, the weighting factor will be  $w_f = \frac{k'}{k}e^{-\delta\tau}$ , where  $k$  and  $k'$  are the total reaction coefficients and  $\delta\tau = \tau' - \tau$  the perturbation in the optical path from the starting point to the collision. It should be noted that, in both cases, for sufficiently small optical path perturbations  $e^{-\delta\tau} \cong 1 - \delta\tau$  can be assumed. Special care must be taken, in any case, over the treatment of the various events during the history tracking, such as, for instance, those connected with the forced collision technique. An estimator of the differential effects of the perturbation upon a given result can be directly set up by collecting, during the history tracking, the differences of interest between the perturbed and the unperturbed environments. The exponential dependence of the weighting fractions causes this instrument to be suitable for very small perturbations, difficult to evaluate using separate calculations.

Obviously, several perturbed environments can be envisaged in the same calculation, each of them being associated with a statistical weight. It must be remarked that, according to this technique, the leading photon history is governed by the reference system, so that the end of its history will cause the end of the histories in the perturbed ones, too. If the  $w_{\min}$  criterion is assumed to end the history tracking, it is convenient, for each perturbed environment, to sum the total unprocessed weighting.

The perturbation technique is foreseen in MOCRA (see below for applications) and in the successive versions of the PREMAR code. A special perturbative technique is also available (Rief, 1984).

To handle statistically significant estimators and to avoid too many contributions being collected, which can cause a loss of precision, batches of assigned numbers of histories can be envisaged. For each batch, the average of the quantity of interest over the processed histories will give the estimator value belonging to that batch. The average value over the number of batches will give the searched estimate. As an example, 100 batches each of 1000 histories could be run. To make it possible to reach a desired precision in the simulation, a restart option over the batches is envisaged, so as to optimize the calculation time.

To facilitate the comparison between two different calculations characterized by small differences in the physical or geometrical parameters, a strategy on the random number generation can be devised for both batches and histories inside a batch, so that each history will begin with its proper initial random number. Unaltered corresponding histories in the two calculations will give in this way the same results for the same quantity. This device is envisaged both in MOCRA and in PREMAR (Premuda et al., 2009; Cupini et al., 2001).

### 3. Simulation of LIDAR and DOAS observations

In general "Remote Sensing" techniques analyze kinds of interactions between a wave and the examined medium. For minor gases in atmosphere electromagnetic radiation is certainly

the most adequate and its interactions with the atmosphere are known as “spectroscopic”. Various spectroscopic techniques have been developed to measure concentrations of atmospheric pollutants.

There are two main types of remote sensing systems: passive remote sensing and active remote sensing. Passive sensors detect natural radiation emitted or reflected by the object or surrounding area being observed. Reflected or scattered sunlight is the most common source of radiation measured by passive sensors. Examples of passive remote sensors include film photography, charge-coupled devices and radiometers. Active collection, on the other hand, emits energy in order to scan objects and areas whereupon a sensor then detects and measures the radiation that is reflected or backscattered from the target. RADAR is an example of active remote sensing where the time delay between emission and return is measured, establishing the location, height, speed and direction of an object.

Remote sensing makes it possible to collect data on dangerous or inaccessible areas. Its applications include monitoring deforestation in areas such as the Amazon Basin, the effects of climate change on glaciers and on Arctic and Antarctic regions, and depth sounding of coastal and ocean depths. Remote sensing also replaces costly and slow data collection on the ground, ensuring in the process that areas or objects are not altered.

Chemical remote sensing systems (radiometers, spectrometers, interferometers) are mounted on many satellites, allowing mapping of pollutants or of minor atmospheric components over large areas of the earth. Orbital platforms thus collect and transmit data from different parts of the electromagnetic spectrum, which in conjunction with aerial or ground-based sensing and analysis, provides researchers with enough information to monitor trends such as El Niño and other natural long- and short-term phenomena. Doppler radar is used, for instance, in enhanced meteorological collection such as wind speed and direction within weather systems. Other types of active collection include plasmas in the ionosphere. Interferometric synthetic aperture radar is used to produce precise digital elevation models of large scale terrain (see RADARSAT, TerraSAR-X, Magellan). Radiometers and photometers are the most common instruments in use, collecting reflected and emitted radiation over a wide range of frequencies. The most common are visible and infrared sensors, followed by microwave, gamma ray and rarely, ultraviolet. They may also be used to detect the emission spectra of various chemicals, providing data on chemical concentrations in the atmosphere.

Among the various remote sensing techniques, LIDAR (Light Detection And Ranging) and DOAS (Differential Optical Absorption Spectroscopy), allowing the analysis of gas and aerosol concentrations in the atmosphere.

The applications of DOAS remote sensing systems are numerous on the ground, and in both airborne and satellite configurations.

LIDAR ground systems are used to detect and measure the concentration of various chemical compounds in the atmosphere, while airborne LIDAR can be used to analyze the ground characteristics, such as the heights of objects and vegetation, more accurately than with radar technology. Underwater LIDAR and LIDAR in a coupled air-ocean system allow the analysis of marine environments in the seas and oceans, for oil spills and, phytoplankton development. In water systems, fluorescence emission and Raman scattering can play a fundamental role.

Below, the Monte Carlo simulation of LIDAR and DOAS observations is discussed, together with proper variance-reducing techniques.

### 3.1 LIDAR systems

A LIDAR system consists of a laser pulsed source at a given wavelength and a receiving telescope collecting the backscattered radiation. From the characteristics of the detected radiation (total intensity, time and spatial distributions) it is possible to obtain information about the nature and concentration of particles or about obstacles encountered along the path. Given an assigned reference system (Oxyz) with z-axis normal to the ground, for simulation purposes, the source S and the telescope T can be schematized as independent plane disks centered at given coordinate points in the geometric environment, each with its own diameter, axis and FOV. In Fig. 3 a graphical schematic 2-D representation is shown. Regarding the telescope, the disk representation can be adequate for atmospheric LIDAR systems, where collision points normally occur far from the telescope disk centre. More realistic descriptions of the telescope device are needed, for instance, in underwater LIDAR systems, as envisaged in PREMAR (Cupini et al., 2001).

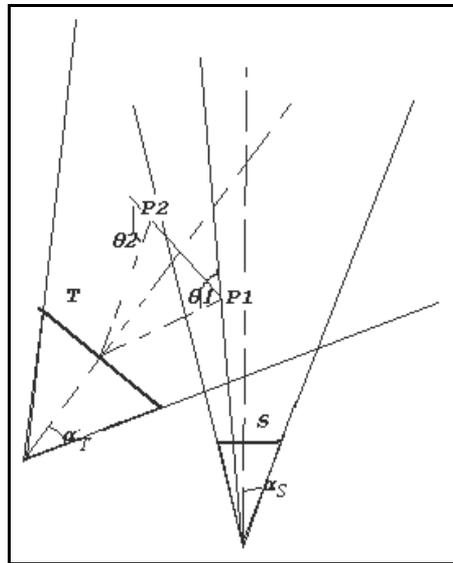


Fig. 3. Schematic disk representation of LIDAR system for simulation purposes. S is the source disk, T is the telescope disk, P1 and P2 are collision points,  $\theta_1$  and  $\theta_2$  are angles between flight direction and direction connecting collision points to the telescope centre.

In Monte Carlo simulation an outgoing ray from the vertex of a fictitious source cone, possibly located outside the atmospheric geometrical system, is uniformly generated, i.e., if  $\alpha_s$  is the angle which characterizes the source aperture angle, an angle cosine with respect to the cone axis is uniformly chosen in  $(\cos\alpha_s, 1)$  with a corresponding azimuth angle uniformly chosen in  $(0, 2\pi)$ . The ray intersection with the source disk identifies the initial coordinates of the source photon, whose motion direction coincides with that of the ray. If the source aperture angle is equal to zero, the photon starting point is sampled uniformly on the disk with its motion direction orthogonal to it. For photons whose collision point falls into the telescope FOV the distance  $R$  between such a point and the centre of the telescope disk is evaluated (see Fig. 3). The expected collision contribution to the intensity  $I_T$  of the radiation collected by the telescope can be written as:

$$I_T = (w_{as}P_a(\Theta)\Delta\Omega + w_{ms}P_m(\Theta)\Delta\Omega)e^{-\tau} \quad (19)$$

where  $w_{as}$  is the product of the current photon weighting and the probability of scattering by aerosol particles,  $w_{ms}$  the analogous quantity for a molecular component,  $P_a(\Theta)$  and  $P_m(\Theta)$  the corresponding phase functions for the angle  $\Theta$  between the flight direction before scattering and the direction from the collision point to the centre of the telescope disk, normalized to 1 over the whole solid angle,  $\tau$  is the optical distance between collision point and the centre of telescope disk. The solid angle element  $\Delta\Omega$  is given by the expression

$$\Delta\Omega = \frac{A_T \cos\Phi}{R^2} \quad (20)$$

where  $A_T$  is the area of the telescope disk and  $\Phi$  the angle between the telescope axis and the disk centre-collision point direction. To avoid infinite variances the calculation can be performed only for  $R$  values greater than a certain threshold. For this purpose, if small distances between the collision point and the telescope disk are of interest, as could happen, for instance, in underwater LIDAR systems, the following  $\Delta\Omega$  formula can be used, where  $r_n$  is the radius of the disk normal to the direction of the photon-to-telescope-disk centre:

$$\Delta\Omega = 2\pi(1 - \cos\theta) = 2\pi \left( 1 - \frac{1}{\sqrt{1 + (r_n/R)^2}} \right) \quad (21)$$

which is reduced to eq. (20) when  $r_n/R \rightarrow 0$ . In Fig. 4 a comparison between the approximate and the correct formulae is given.

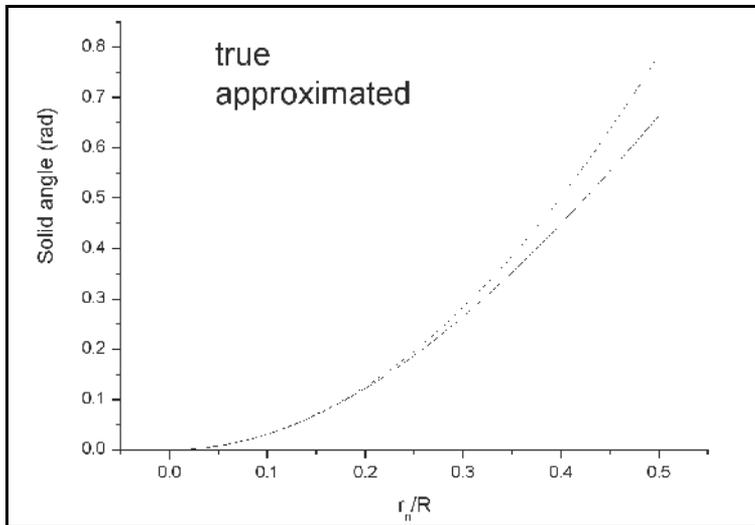


Fig. 4. Comparison between exact and approximated solid angle expressions.

The aerosol  $I_A$  and molecular  $I_M$  contributions to the intensity  $I_T$  of radiation in eq. (19) can be considered separately in the calculation. According to assigned time intervals, the time elapsed distribution from the emission at the source to detection at the telescope can be

obtained. Along the axis of the telescope disk, starting from its surface plane, spatial intervals can be considered. Spatial distributions of the radiation intensities on the telescope can thus be obtained by projecting onto its axis the collision points within the FOV. Within this context, a possibly interesting quantity is the spatial distribution of the so-called backscattering ratio between the total and molecular radiation intensities.

A comparison between the Monte Carlo results for the backscattering ratio obtained with PREMAR and experimental ones obtained at the ENEA centre at Brasimone lake (Bologna), located at an altitude of 0.91 Km, is given in Fig. 5. The scheme of the vertical Lidar system is given in Fig. 6 (not to scale), the disk source having a diameter  $d_1 = 2.8$  cm with a FOV of  $\alpha_1 = 0.12$  mrad and a telescope disk diameter  $d_2 = 80$  cm with a FOV = 0.35 mrad, with a distance of 80 cm between the two disk centers. A wavelength of  $0.532 \mu\text{m}$  was considered. The measurements were carried out over two periods: the first in September 1993, to verify the presence of volcanic aerosols following the eruption of Pinatubo, which occurred in June 1991; the second in February 1995. As can be seen, a satisfactory agreement is obtained when using the high volcanic profile foreseen by MODTRAN for the fall-winter season (on the left) and background stratospheric aerosol profile, again predicted by MODTRAN for fall-winter (on the right).

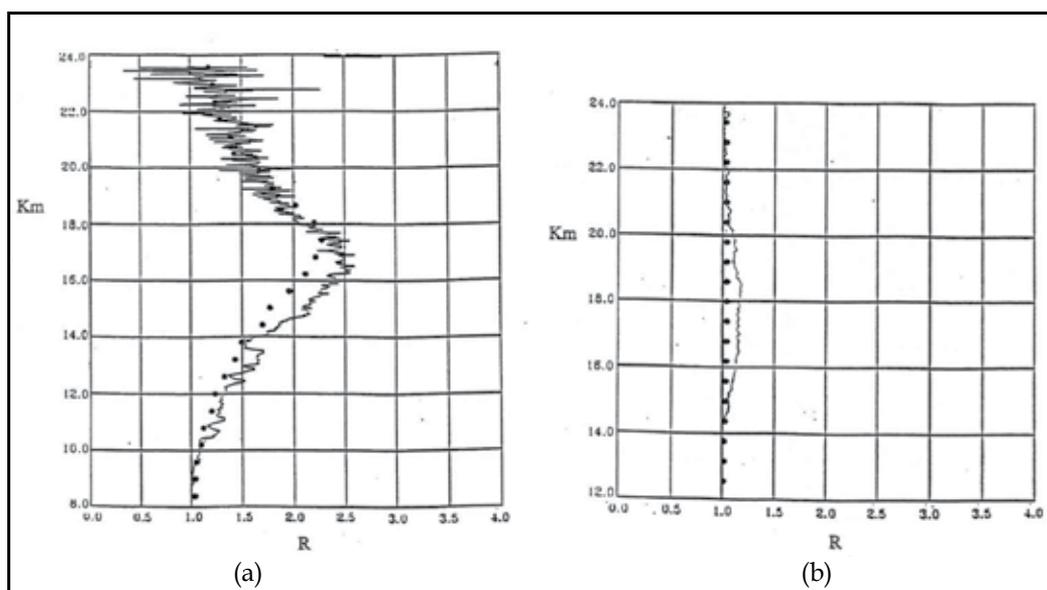


Fig. 5. Backscattering ratio  $R$ : comparison between PREMAR simulation and measurements carried out with LIDAR at the Brasimone ENEA centre in September 1993, after the eruption of Pinatubo, which occurred in June 1991, and in February 1995. The simulation was performed using: a) a high volcanic MODTRAN profile for fall-winter ; b) a background stratospheric aerosol MODTRAN profile for fall-winter (Cupini et al. 1997).

To reduce the variance of the calculation, besides the forced collision technique previously described, the local forced collision and the splitting techniques can be used when the contribution to the backscattered radiation of particular layers with small optical depths is of special interest.

According to the local forced collision technique, when a photon particle crosses a boundary of the layer of interest, the photon coordinates of the crossing point are memorized and a virtual collision is forced to occur in the layer with the proper statistical weight for the photon. The foreseen statistics related to this collision point are performed and the memorized coordinates of the photon redefined. A path is then chosen in the usual way, without further statistics if a collision occurs within the layer. In this latter case, the device can be repeated as in the case of the crossing of the layer boundary (see Fig. 7). The advantage of such a technique resides in the fact that efforts are concentrated only upon the layers of interest and, moreover, the photon travelling weight, unlike the general forced collision technique, remains unchanged.

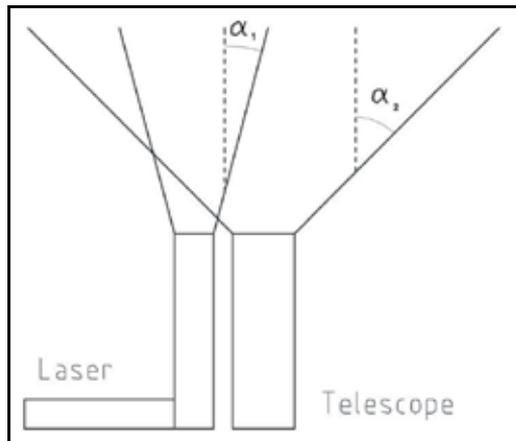


Fig. 6. Schematic representation of the vertical LIDAR system at the Brasimone ENEA Centre:  $\alpha_1 = 0.12$  mrad,  $\alpha_2 = 0.35$  mrad

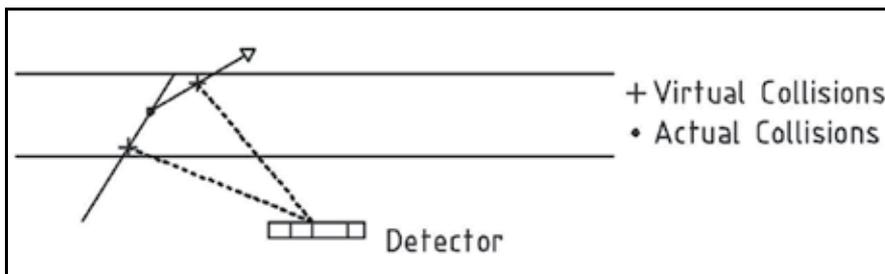


Fig. 7. Local forced collision: the photon undergoes two virtual collisions and one actual collision within the layer of interest. Only the virtual collisions give rise to contributions to the signal collected by the detector.

Regarding the splitting technique, when a photon crosses a boundary of the layer of interest with an assigned splitting index  $m$ , analogously to a local forced collision, the photon coordinates of the crossing point are memorized. From this point a number  $m$  of photons are generated, each with a weight  $w_i = w/m$  ( $i=1, \dots, m$ ). All the photons are processed within the layer, selecting their starting path according to the exponential distribution, ending their history, except for the last one, if they leave it. If the last photon emerges from the layer, its

weight is multiplied by the splitting index  $m$  and its history tracking continues (see Fig. 8). If more adjacent layers with the same splitting index  $m$  are simultaneously considered, the processing of the generated photon can continue into the new layer with the same splitting index. Adjacent layers with different splitting indices can be considered, as foreseen in PREMAR but, for simplicity, are not described here. Excessively high values of  $m$  should be avoided: a value of between 2 and 10 should be adequate. For special treatment of the splitting technique, see, for instance, Burn (Burn, 1995; Burn 1997). It is of interest to underline that the local forced collision and the splitting technique can be adopted within the same calculation for the same or different layers.

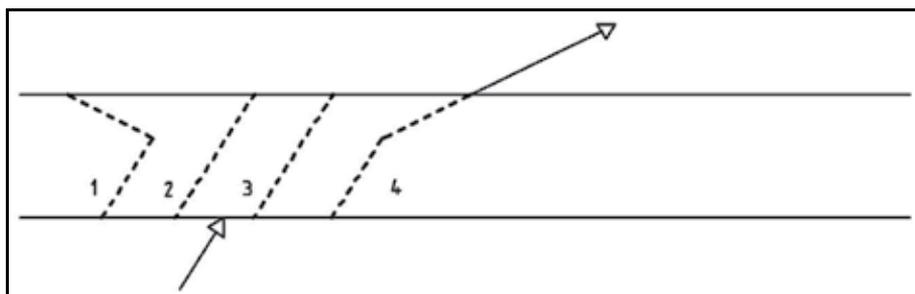


Fig. 8. Splitting technique for a layer with splitting index  $m=4$ . The first photon undergoes a collision and its history ends when it hits the layer boundary. The second and the third photon would leave the layer without colliding but their histories end. Only the history of the last generated photon continues, as it emerges from the layer after a collision.

### 3.2 DOAS systems

DOAS is a remote sensing method which identifies and quantifies the trace gases in the atmosphere taking advantage of their absorption structures in the near UV and visible wavelength range. The molecular absorption is analyzed to obtain the concentration of the trace gases integrated along the optical path between the source and the receiver (Slant Column Density, SCD). After being introduced by Noxon (1975) during stratospheric trace gas studies at Fritz Peak observatory in Colorado, DOAS quickly became one of the most promising methods for determining the role of minor gases affecting ozone depletion in the lower stratosphere. In the late 1970s, Platt and Perner carried out DOAS measurements of CHO<sub>2</sub>, O<sub>3</sub>, and NO<sub>2</sub> in maritime and rural areas of Northern Europe (Platt et al., 1979; Platt & Perner, 1980). During the following decades DOAS techniques have been applied to urban pollution analysis and monitoring as well as in the climatic and environmental fields following two different methodological and instrumental approaches: 1) the active mode which can, by using an artificial source of radiation, perform measurements of atmospheric minor gas concentrations, integrated along the optical path between the lamp and the receiving telescope of the instrument (Perner et al., 1976; Evangelisti et al., 1995; Stutz & Platt, 1997a,b). 2) The passive mode, using diffuse solar radiation as the radiation source, carries out measurements of the examined gas column content along both the *vertical* (zenith-sky) and *oblique* (also called *Off-Axis* or MAX-DOAS (Wagner et al. 2007a)) directions. The zenith-sky configuration is especially used for observations in the lower stratosphere region for research mainly related to climate studies; the Off-Axis mode is

mainly applied in environmental monitoring (Petrifoli et al., 2002; Hönninger et al., 2004; Bortoli et al. 2005, Giovanelli et al., 2006).

Particularly in the passive mode, both in vertical and oblique configurations, an Air Mass Factor (AMF) is calculated as a fundamental part of the interpretation of DOAS data in atmospheric observations using a Radiative Transmission Model (RTM) which also requires the use of a multiple-scattering configuration to describe the radiation passage through the atmosphere.

The AMF concept can be generalized to the recently developed "off-axis" configuration (Hönninger et al., 2004) and appropriate RTMs have to be used for data interpretation.

Another recent application of this type of remote sensing system is the so-called ToTaL-DOAS (Topographic Target Light scattering DOAS) which is a novel experimental procedure to retrieve trace gas concentrations present in the lower atmosphere. Scattered sunlight reflected from natural or artificial targets located at different distances are analyzed to retrieve the spatial distribution of the concentration of different trace gases such as NO<sub>2</sub>, SO<sub>2</sub> and others (Frins et al., 2006; Frins et al., 2008; Louban et al., 2008). In this case RTMs are required to compute the Equivalent Path Length (EPL), i.e. the distance of a fictitious radiation source giving the same signal collected by the detector (Premuda et al., 2009).

For the simulation of radiometer or DOAS devices, a reference system O(x,y,z) can be assumed with its origin at the Earth's centre and the z-axis normal to ground surface. Direction cosines  $u_s, v_s, w_s$  will define the solar source radiation direction. An external atmospheric boundary surface is considered, over which the starting points of the solar radiation are projected. Two kinds of detectors can be taken into account: the first being schematized as a point detector located on the ground, looking upward, and the second as a disk detector of a given radius, with its central point geometrical co-ordinates  $(x_0, y_0, z_0)$  and normal viewing direction  $(u_0, v_0, w_0)$ . In the latter case, the FOV is assumed. The disk receiver can be reduced to a point with an assigned line-of-sight  $(u_0, v_0, w_0)$ .

In the first case, if  $I(z)$  is the radiance intensity due to a solar ray reaching the detector from a height  $z$ , the total intensity  $I_T$  can be obtained by integrating the contributions from all values of  $z$  from the ground to the upper boundary of the atmosphere:

$$I_T = \int I(z) dz . \quad (22)$$

If the entire vertical range is subdivided into N intervals of thickness  $\Delta z_i$  ( $i=1, \dots, N$ ) it can equivalently be written as

$$I_T = \sum_i \Delta z_i \int_{\Delta z_i} I(z) (dz / \Delta z_i) . \quad (23)$$

In the Monte Carlo simulation, a stratified sampling procedure can be performed, as in MOCRA, which consists of randomly choosing a set of  $z$  points uniformly distributed within each height interval  $\Delta z_i$ , and estimating the corresponding  $I(z)$  value. The average intensity value over the points belonging to the  $i$ -th interval times the vertical layer thickness  $\Delta z_i$  gives the contribution of that interval to the total intensity. The single scattering radiance  $I_S$  is calculated using a forward Monte Carlo simulation of the photon path from the solar source coordinate on the external geometrical boundary, computed on the basis of the  $z$  altitude of each selected point and of the assigned solar zenith angle,

taking into account the refraction when the photon path crosses a boundary between geometrical regions with different refractive indices, until the z-axis is reached. The different arrival points on the z-axis will be utilized to compute the contribution to  $I_S$  of the appropriate intervals. The average geometrical and optical slant and vertical paths can be evaluated. The multiple scattering radiance  $I_M$  is calculated using a backward Monte Carlo simulation from the selected vertical points to the sun, as previously described. The contribution of each interval to  $I_M$  is computed and the total scattering radiance  $I_T = I_S + I_M$  evaluated. In the  $I_M$  calculation, albedo and refraction phenomena can be envisaged, possibly excluding, as in MOCRA, refraction along the path from the last scattering point to the sun. Perliski and Soloman (Perliski & Solomon, 1993) carried out a backward Monte Carlo simulation in a spherical shell model atmosphere to calculate the air mass factor for a vertical upward-looking detector.

In case of disk detectors with assigned diameter, FOV, position and orientation, let  $P_d = (x_d, y_d, z_d)$  be the centre of the disk,  $D (\geq 0)$  its diameter and  $(u_d, v_d, w_d)$  the direction cosines of the outgoing direction normal to the disk. Moreover, let  $\alpha (> 0)$  be the field of view of the detector.

By assuming  $D > 0$ , a fictitious cone with a semi-amplitude  $\alpha$  and the disk as its base can thus be generated. The vertex point  $P_v \equiv (x_v, y_v, z_v)$  (possibly located outside the atmospheric geometrical system as previously mentioned for the LIDAR source) can be considered as the virtual source point in a backward Monte Carlo simulation, with initial motion directions uniformly distributed within the cone. For each  $\Omega$  direction, the crossing point with the detector disk will give the true starting point for the particle. Let  $\Delta\Omega$  be the solid angle amplitude and  $I(\Omega)$  the radiance intensity on the receiver from the  $\Omega$  direction. The total intensity  $I_T$  over all possible directions is assumed to be:

$$\begin{aligned} I_T &= \int_{\Delta\Omega} I(\Omega) d\Omega \\ &= \Delta\Omega \int_{\Delta\Omega} I(\Omega) d\Omega / \Delta\Omega \end{aligned} \quad (24)$$

Eq. (24) can be simulated through a standard backward Monte Carlo estimating procedure for  $I_T$ .

If  $D = 0$ , analogous considerations can be made, with the true starting point coinciding with the virtual one. If  $\alpha = 0$ , the starting point is chosen uniformly on the disk and the initial direction is that of the detector axis.

The PROMSAR (PROcessing of Multi - Scattered Atmospheric Radiation) code (Palazzi et al. 2005) was developed to perform backward Monte Carlo simulations of DOAS observations for vertical and off-axis looking detectors. In MOCRA, through a 3D multi-region geometry, optionally selectable by the user, the possibility of defining different topographic and atmospheric scenarios for a set of user-defined regions is included. This allows simulating, for instance, a horizontal point detector with a line of sight perpendicular to a vertical obstacle 100 m high and 50 m wide, at a distance of 1 Km from the detector. The solar direction cosines were obtained by means of the astronomical parameters. Fig. 9 shows a comparison of the simulation results with measurements obtained from the DOAS spectrometer TropoGAS (Tropospheric Gas Analyzer Spectrometer) working, in this case, in a horizontal-view configuration at a distance of 1 Km from a house (Premuda et al., 2009).

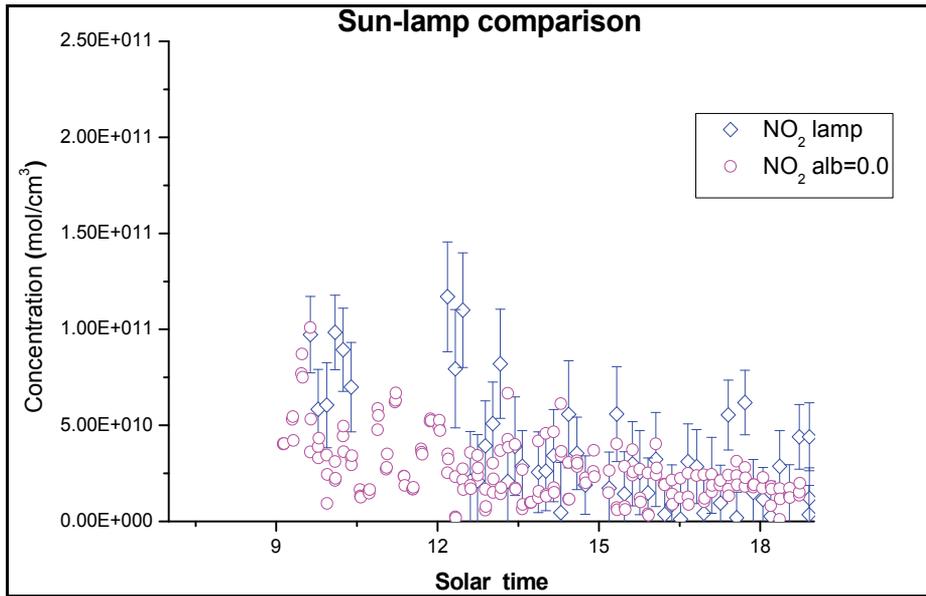


Fig. 9. Passive-active measurements comparison at S. Pietro Capofiume (11.6226° E, 44.6538° N), Bologna, Italy. Measurements were performed almost simultaneously by the same instrument in active (Xenon lamp) and passive modes. The lamp was placed in front of a house at 1 Km from the spectrometer. A passive measurement towards the house was inserted in the measurement table. Only data with errors of less than 30% are plotted. The axis covers the time interval 7-19 (hours) with a linear scale.

Data obtained from passive measurements by means of the simulation show a very good agreement with the simultaneous active measurements used as a reference for the comparison.

As for the AMF calculation, let  $I$  and  $I^*$  be the radiance (single or total scattering) detected by the receiver with and without the trace gas of interest, respectively. The AMF of the species is defined as (Sarkissian et al., 1995):

$$AMF = -\frac{\ln\left(\frac{I}{I^*}\right)}{\delta_a} \quad (25)$$

where  $\delta_a$  is the vertical absorption optical depth of the molecular species of interest. In the case of weak absorption, an approximate formula for AMF is given by :

$$AMF = \frac{\delta_{OPT}}{\delta_a} \quad (26)$$

where  $\delta_{OPT}$  is the intensity-weighted absorption optical path over all the collisions or reflections which contribute to the total intensity  $I^*$ . If  $I \approx I^*$  can be assumed, the calculation can be performed directly within the reference system containing the trace gas (Sarkissian et al., 1995). In MOCRA, when calculating  $\delta_{OPT}$ , perturbed radiance values are used if perturbative calculations are performed, whereas unperturbed radiance values are

used if this is not the case. Due to the possibility of perturbative calculations, an estimation of the correct air mass factor can be carried out considering a perturbed environment without the absorption contribution of the molecular particle of interest. Fig. 10 shows plots of exact and approximate AMFs for single-scattering radiance, computed by MOCRA for  $O_3$  (Fig. 10.a) and  $NO_2$  (Fig. 10.b) at wavelength of 310 nm for a mid-latitude summer atmospheric environment with urban extinction. It can be seen that for ozone there is a significant difference between the exact and approximate value at higher solar zenith angles, due to its marked absorption at the given wavelength enhanced by the greater path length, whereas the values for  $NO_2$ , which is a weak absorber, are almost identical to each other.

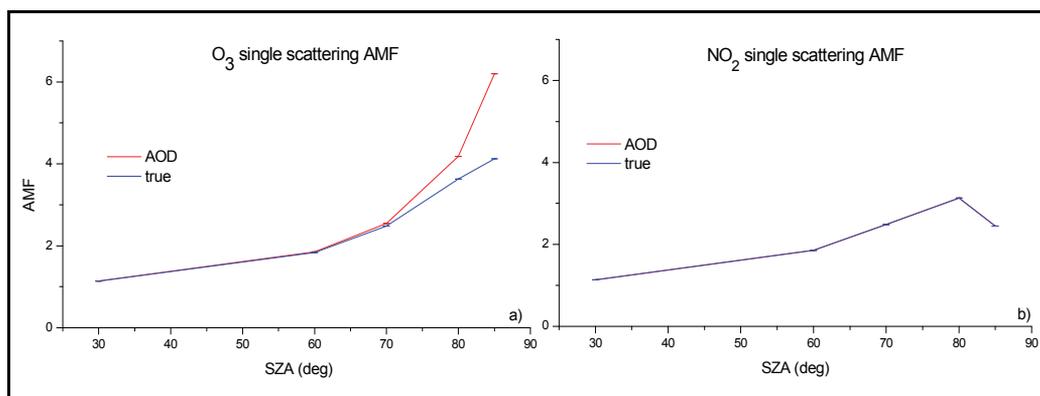


Fig. 10. AMF formula comparison for a mid-latitude summer atmosphere,  $\lambda=310$  nm, urban extinction vis.=35 Km: a)  $O_3$  single-scattering radiance AMF; b)  $NO_2$  single-scattering radiance AMF.

To generate the gas profiles, starting from the gas SCDs via an inversion method, the “box air mass factors” which describe the sensitivity of the measurements as a function of atmospheric layer altitude have to be computed. They are defined as (Pukite et al., 2006):

$$AMF_b = \frac{dSCD_g}{dVCD_b} = -\frac{1}{h_b \sigma_b} \frac{d \ln I_g}{dn_b} = \frac{1}{h_b} \frac{d \ln I_g}{d\beta_b}. \quad (27)$$

where  $n_b$  is the number density of the gas in the box,  $h_b$  is the box height and  $\beta_b$  is the absorption coefficient,  $VCD_b$  (Vertical Column Density) is the integral of concentration along the vertical direction within the box. This is computed, for instance in the TRACY (Trace gas RAdiative Transfer Monte Carlo Y(I)mplementation) series models (Pukite et al., 2006; Kühl et al., 2008), simulating radiative transfer, using a backward Monte Carlo scheme.

A Monte Carlo evaluation of the box AMF can be performed as previously described for the AMF over the entire atmospheric environment, by considering the box with and without the trace gas of interest and assuming the optical box height as  $\delta_a$  in eq. (25) and eq. (26). In eq. (26) the absorption optical distances which effectively contribute to the box AMF will be those belonging to the box. In MOCRA several boxes can be considered simultaneously, each consisting of one or more vertical layers.

PROMSAR Box Air Mass Factors and radiances were validated through comparison with a series of state-of-the-art UV/visible RTMs (Wagner et al., 2007b).

To optimize the type of calculation which involve single- and multiple-scattering radiances being computed together with, where envisaged, the reflected radiance, simulation techniques can be adopted which force the path starting from the source to undergo the desired event. In the case of single- and multiple-scattering radiances, for the photon, born at the source with an initial weight  $w = w_0$ , the probability  $p_c$  of collision within the system is computed. According to this probability, a photon with a starting weight  $w = w_0 p_c$  is forced to collide, thus giving its contribution to the desired radiances. If the reflected contribution from the ground to the detector is not required, the simulation can stop and a new source photon can be chosen. Otherwise, if the initially selected motion direction crosses the ground, a new photon, with the same characteristics of the previous one, starts from the source with a weight  $w = w_0(1.0 - p_c)$ , which will be reflected according to the envisaged albedo coefficient. In the case where only the reflected contribution is of interest, this last simulation is the only one performed. It should be noted that, if scattered and reflected contributions are both required, each history is split into two sub-histories for a given source-starting photon. Such a technique can efficiently be applied when there are low probabilities that the starting photon will collide or be reflected. In MOCRA the standard simulation procedures together with those described here are envisaged. It should be stressed that this technique is independent from the general, previously described forced collision technique, which can not be applied when the direction of motion crosses the ground. The forced collision technique can in any case be applied for collisions following the first one or reflection.

Clearly, such a device could also be adopted for Lidar system simulations, such as when ground characteristics are to be examined using airborne or satellite Lidar systems.

#### 4. Possible future developments

The simulation methods presented here are essentially devoted to the solution of RTE in the UV-visible spectral regions, where the internal source term  $S(r)$  in Eq. (2) can be neglected. But remote sensing problems, such as MIPAS (Michelson Interferometer for Passive Atmospheric Sounding) data analysis, dealing with the infrared spectral region, requires a consideration of the presence of sources at the Earth's surface and within the atmospheric system, so that  $S(r)$  cannot be neglected. In fact, the atmosphere presents a wide absorption window in the visible frequencies, whereas it is a strong absorber in the infrared, and, according to Kirchoff's law, the emission and absorption coefficients for each wavelength are the same. The principal problem to be solved is thus how to represent these two kinds of sources.

To a first approximation, the external source at the Earth's surface should be set equal to Planck's blackbody function  $B(\nu, T)$

$$B(\nu, T) = \frac{2h\nu^3}{c^2} \frac{1}{\frac{h\nu}{e^{KT}} - 1} . \quad (28)$$

In general, the internal source and the scattering and absorption macroscopic cross-sections depend upon the description of atoms and molecules which constitute matter and their energetic status and there is no simple relationship between them.

One frequently advanced hypothesis is that of Local Thermodynamic Equilibrium (LTE) (Pomraning, 1982). This assumes that medium properties are dominated by atomic collisions and at each time at each point the system is in thermodynamic equilibrium and the radiation field, even though very different from  $B(\nu, T)$ , does not influence thermodynamic equilibrium. If this hypothesis holds, at each point of the system and at each time  $t$ , only two variables (besides the system composition) are required to calculate the internal source  $S$  and the absorption and scattering cross-sections  $\sigma_a$  and  $\sigma_s$ . Thus, with the usual meaning of the symbols,

$$k'_a = k_a(\nu) [1 - \exp(-h\nu / KT)] \quad (29)$$

and the internal source can be defined as

$$S(r) = k'_a(\nu) B. \quad (30)$$

On the basis of these considerations a further development of the work done within this framework could be the extension of MOCRA to the infrared spectral range introducing a source at the Earth's surface defined by Eq. (28) and an internal radiation source defined by Eq. (30) according to surface or air temperature and radiation frequency. Blackbody internal sources are foreseen in MODTRAN codes.

More detailed studies should be devoted to molecular scattering simulation, taking into account, besides the Rayleigh scattering, Raman scattering and their mutual connections (Young, 1981). In PREMAR, Rayleigh and Raman scattering are considered separately for water environments.

## 5. References

- Acharya, P., Adler-Golden, S. M., Anderson, G. P., Berk, A., Bernstein, L. S., Chetwynd, J. H., et al., (1998). *Modtran version 3.7/4.0 user's manual*, Air Force Research Laboratory, Space Vehicles Directorate, Air Force MATERIEL Command, Hanscom AFB, MA.
- Berk, A., Bernstein, L. S. & Robertson, D. C. (1989). *MODTRAN: a moderate resolution model for LOWTRAN7*, Geophysics Laboratory, Air Force Systems Command, United States Air Force, Hanscom AFB, MA
- Berk, A., Anderson, G. P., Acharya, P. K., Chetwynd, J. H., Bernstein, L., Shettle, E. P. et al., (1999). *MODTRAN4 user's manual*, Air Force Research Laboratory, Space Vehicles Directorate, Air Force Materiel Command, Hanscom AFB, MA
- Bortoli, D., Giovanelli, G., Ravegnani, F., Kostadinov, I. & Petritoli, A., (2005). Stratospheric Nitrogen Dioxide in the Antarctic, *Int J. Of Remote Sensing*, Vol. 26, 16, 3395-3412.
- Burn, K. W., (1995). Extending the Direct Statistical Approach to Include Particle Bifurcation between the Splitting Surfaces, *Nucl. Sci. Eng.*, 119, 24.
- Burn, K. W., (1997). A New Weight Depending Statistical Approach Model, *Nucl. Sci. Eng.*, 125,128.

- Clough, S. A., Kneizys, F. X., Rothman, L. S., Gallery, W. O. (1981). Atmospheric Spectral Transmittance and Radiance: FASCOD1B, *Proceedings of SPIE* Vol. 277 Atmospheric Transmission
- Coletti, A. & Fiocco, G. (1980). Monte Carlo Simulation of a Pulsed Laser Beam Diffusing through Fog : Spatial and Temporal Structure of the Echoes, *Il Nuovo Cimento*, Vol. 3C, No. 6.
- Collins, D. G., Blättner, W. G., Wells, M. B. & Horak, H. G. (1972). Backward Monte Carlo calculations of the polarization characteristics of the radiation emerging from spherical-shell atmospheres, *Appl. Opt.*, Vol. 11, 2684-2696.
- Cupini, E., Borgia, M. G., & Premuda, M. (1997), *Il codice PREMAR per la simulazione Montecarlo del trasporto della radiazione nell'atmosfera*. (RT/INN/97/5).
- Cupini, E., Ferro, G. & Ferrari, N., (2001) *Monte Carlo Analysis of Radiative Transport in Oceanographic Lidar Measurements*. (ENEA/RT/INN/2001/7).
- Cupini, E., Ferro, G. & Sukhanov, A. (2005). Introduction of 3-dimensional atmospheric radiative transport in the PREMAR Monte Carlo code, *Proceedings of XII-th Joint International Symposium "Atmospheric and Ocean Optics, Atmospheric physics"*, Tomsk (2005)
- De Matteis, A. & Simonini, R. (1978a). A New Monte Carlo Approach to the Adjoint Boltzmann Equation, *Nuclear Science and Engineering*, Vol. 65, 93-105
- De Matteis, A. & Simonini, R. (1978b). A Monte Carlo Biasing Scheme for the Adjoint Photon Transport, *Nuclear Science and Engineering*, Vol. 67, 309-316
- De Matteis, A. & Pagnutti, S. (1988). Parallelization of Random Number Generators and Long-range Correlations, *Numer. Math*, 53, 595-608.
- Edlén, B., (1966). The Refractive Index of Air, *Metrologia*, Vol. 2, No. 2, 71-80.
- Evangelisti, F., Baroncelli, A., Bonasoni, P., Giovanelli, G., & Ravegnani, F., (1995). Differential optical absorption spectrometer for measurement of tropospheric pollutants, *Applied optics*, Vol. 34, No. 15, 2737-2744.
- Frins, E., Bobrowski, N., Platt, U. & Wagner, T., (2006). Tomographic multi-axis-differential optical absorption spectroscopy observations of Sun-illuminated targets: a technique providing well-defined absorption paths in the boundary layer, *Applied Optics*, Vol. 45, N. 24, 6227-6240,
- Frins, E., Platt, U. & Wagner, T., (2008). High spatial resolution measurements of NO<sub>2</sub> applying Topographic Target Light scattering-Differential Optical Absorption Spectroscopy (ToTaL-DOAS), *Atmos. Chem. Phys.*, 8, 7595-7601.
- Giovanelli, G., Palazzi, E., Petritoli, A., Bortoli, D., Kostadinov, I., Margelli, F., Pagnutti, S., Premuda, M., Ravegnani, F., & Trivellone, G., (2006). Perspectives of 2D and 3D mapping of atmospheric pollutants over urban areas by means of airborne DOAS spectrometers. *Annals of Geophysics*, Vol. 49, N.1, 133-142.
- Goody, R. M. & Yung, Y. L. (1995). *Atmospheric Radiation: Theoretical Basis*, Oxford University Press, ISBN 0 19 510291 6, New York.
- Hönninger, G., Friedeburg, C.V., & Platt, U., (2004). Multi axis differential absorption spectroscopy (MAX-DOAS). *Atmos. Chem. Phys.*, Vol. 4, 231-254
- Kneizys, F. X., Clough, S. A., Shettle, E. P. (1983a). Atmospheric Attenuation of Laser Radiation, *Proceedings of SPIE*, Vol. 410 Atmospheric Transmission

- Kneizys, F. X., Shettle, E. P., Gallery, W. O., Chetwynd, J. H. Jr, Abreu, L. W., Selby, J. E. A., Clough, S. A. & Fenn, R. W., (1983b). *Atmospheric Transmittance/Radiance : Computer Code LOWTRAN 6*, AFGL-TR-83-0187.
- Kneizys, F. X., Clough, S. A., Shettle, E. P., Rothman, L. S., Fenn, R. W. (1984). *Linear Absorption and Scattering of Laser Beams*, Air Force Geophysical Laboratory, AFGL-TR-84 0265.
- Kneizys, F., Shettle, E. P., Abreu, L. W., Chetwynd, J. H., Anderson, G. P., Gallery, W. O., Selby, J. E. A. & Clough, S. A., (1998). *User Guide to LOWTRAN-7*, Air Force Geophysics Lab., Hanscom AFB, AFGL-TR 880177.
- Kneizys, F., Robertson, D., Abreu, L. W., Acharya, P., Anderson, G. P., Rothman, L. S. et al. (1996)., *The MODTRAN 2/3 Report and LOWTRAN7 MODEL*, Phillips Laboratory, Geophysics Directorate, Hanscom AFB, MA
- Knuth, D. E. (1981). *The Art of Computer Programming*, Vol.2: Seminumerical Algorithms, 2nd ed., Addison-Wesley, Reading, MA.
- Kondratyev, K. Ya. (1969). *Radiation in the Atmosphere*, Academic Press, New York San Francisco London
- Kühl, S., Puķite, J., Deutschmann, T., Platt, U., & Wagner, T., (2008). SCIAMACHY Limb Measurements of NO<sub>2</sub>, BrO and OClO, Retrieval of vertical profiles: Algorithm, first results, sensitivity and comparison studies, *Adv. Sp. Res.*, Vol. 42 (10), 1747-1764.
- Lenoble, J. (Editor) (1977). *Standard Procedures to Compute Atmospheric Radiative Transfer in a Scattering Atmosphere - Volume I*, International Association of Meteorology and Atmospheric Physics (IAMAP), Boulder, Colorado, USA
- Liboff, R. L. (1989). *Kinetic Theory: Classical, Quantum and Relativistic Descriptions*, Prentice Hall, Englewood Cliffs New Jersey
- Liou, K. N., (1998). *Radiation and Cloud Processes in the Atmosphere: Theory, Observation and Modeling*, Oxford University Press, ISBN 9780195049107, New York
- Louban, I., Piriz, G., Platt, U. & Frins, E., (2008). Measurement of SO<sub>2</sub> and NO<sub>2</sub> applying ToTaL-DOAS from a remote site, *J. Opt. A: Pure Appl. Opt.*, 10 104017, doi:10.1088/1464-4258/10/10/104017.
- Marseguerra, M. & Zio, E. (2002). *Basics of the Monte Carlo Method with Application to System Reliability*, LiLoLe-Verlag GmbH (Publ. Co. Ltd.), ISBN 3-934447-06-6, Hagen, Germany.
- Mayer, B. & A. Kylling (2000). Three-dimensional radiative transfer calculations with the MYSTIC model. *Poster presentation at IRS 2000*, St. Petersburg, Russia, 24 - 29 July 2000.
- Noxon, J. F. (1975). Nitrogen dioxide in the stratosphere and troposphere measured by ground-based absorption spectroscopy, *Science*, 189, 547-549.
- Pace, G., Cacciani, M., Di Sarra, A., Fiocco, G. & Fua`, D. (2003). Lidar observations of equatorial cirrus clouds at Mahe` Seychelles, *J. Geophys. Res.*, Vol. 108, N. D8, 4236
- Palazzi, E., Petritoli, A., Giovanelli, G., Kostadinov, I., Bortoli, D., Ravegnani, F., Sackey, S. S., (2005). PROMSAR: A backward Monte Carlo spherical RTM for the analysis of DOAS remote sensing measurements, *Adv. Space Res.*, Vol. 36, N.5, 1007-1014.
- Palazzi, E. (2008). Retrieval of trace gases vertical profile in the lower atmosphere combining Differential Optical Absorption Spectroscopy with radiative transfer models,

- Bologna University, Italy, PhD thesis, available at: [http://amsdottorato.cib.unibo.it/983/1/Tesi\\_Palazzi\\_Elisa.pdf](http://amsdottorato.cib.unibo.it/983/1/Tesi_Palazzi_Elisa.pdf)
- Perliski, L. M., & Solomon, S. (1993). On the Evaluation of Air Mass Factors for Atmospheric Near-Ultraviolet and Visible Absorption Spectroscopy, *J. Geophys. Res.*, 10363-10374
- Perner, D., Ehhalt, D. H., Patz, H. W., Platt, U., Roth, E. P., & Volz, A., (1976). OH-radicals in the lower troposphere, *Geophys. Res. Lett.*, Vol 3, 466-468
- Petricoli, A., Giovanelli, G., Ravegnani, F., Kostadinov, Iv., Bortoli, D., & Oulanovsky, A., (2002). Off-axis measurements of atmospheric trace gases from airborne UV/Vis spectrometer, *Applied Optics*, Vol. 41, 5593-5599.
- Platt, U., Perner, D. & Patz, H. W., (1979). Simultaneous measurement of atmospheric CH<sub>2</sub>O, O<sub>3</sub> e NO<sub>2</sub> by differential optical absorption, *J. Geophys. Res.*, Vol. 84, 6329-6335.
- Platt, U. & Perner, D., (1980). Direct measurement of atmospheric CH<sub>2</sub>O, HNO<sub>2</sub>, O<sub>3</sub>, NO<sub>2</sub>, and SO<sub>2</sub> by differential optical absorption in the near UV, *J. Geophys. Res.*, Vol. 85, 7453-7458
- Platt, U., Stutz, J., (2008). *Differential Optical Absorption Spectroscopy. Principles and applications*, Springer-Verlag Berlin Heidelberg.
- Pomraning, G. C. (1982). *Radiation Hydrodynamics*, Los Alamos National Laboratory Radiation Hydrodynamics Short Course Attendees, available at: <http://www.osti.gov/energycitations/servlets/purl/656708-zO5SF0/webviewable/656708.pdf>
- Premuda, F. & Palestini, A. (1982). Diffusive Formulations Arising from Kinetics of Charged Test Particles in a Uniform Field, *Jour. Appl. Math. And Phys. (ZAMP)*, Vol. 33, 783.
- Premuda, M. (1994). *Modellistica fisica e matematica del trasporto radiativo nell'atmosfera e simulazione Montecarlo*, Degree Thesis, in italian.
- Premuda, M., Masieri, S., Bortoli, D., Margelli, F., Ravegnani, F., Petricoli, A., Kostadinov, I., Giovanelli, G. & Cupini, E. (2009). A Monte Carlo simulation of radiative transfer in the atmosphere applied to ToTaL-DOAS, In: *Remote Sensing of Clouds and the Atmosphere XIV*, Richard H. Picard, Klaus Schäfer, Adolfo Comerón, Evgueni I. Kassianov, Christopher J. Mertens (Eds.), *Proceedings of SPIE* Vol. 7475 (SPIE, Bellingham, WA 2009) 74751A.
- Puķite, J., Kūhl, S., Deutschmann, T., Wilms-Grabe, W., Friedeburg, C., Platt, U., & Wagner, T., (2006). Retrieval of stratospheric trace gases from SCIAMACHY limb measurements, *Proceedings of the First Atmospheric Science Conference, 8-12 May, ESA/ESRIN, Frascati, Italy, ESA SP-628*.
- Puķite, J., Kūhl, S., Deutschmann, T., Platt, U., & Wagner, T., (2008). Accounting for the effect of horizontal gradients in limb measurements of scattered sunlight, *Atmos. Chem. Phys.*, Vol. 8, 3045-3060.
- Richtsmeier S. & Sundberg, R. (2009). Full Spectrum Broken Cloud Scene Simulation, In *Remote Sensing of Clouds and the Atmosphere XIV*, Richard H. Picard, Klaus Schäfer, Adolfo Comerón, Evgueni I. Kassianov, Christopher J. Mertens (Eds.), *Proceedings of SPIE* Vol. 7475 (SPIE, Bellingham, WA 2009) 7475-19.

- Rief, H. (1984). Generalized Monte Carlo Perturbation Algorithms for Correlated Sampling and a Second-Order Taylor Series Approach, *Ann. Nucl. Energy*, Vol. 11, No 9, 455
- Roscoe, H. K., Johnston, P. V., Van Roozendal, M., Richter, A., Roscoe, J., Lambert, J-C., Hermans, C., DeCuyper, W., Dzienus, S., Winterrath, T., Barrows, J., Sarkissian, A., Goutail, F., Pommereau, J-P., D'Almeida, E., Hottier, J., Coureul, C., Didier, R., Pound, I., Barlet, L. M., McElroy, C. T., Kerr, J. E., Elokhov, A., Giovanelli, G., Ravegnani, F., Premuda, M., Kostadinov, I., Erle, F., Wagner, T., Pfeister, K., Kenntner, M., Marquard, L. C., Gil, M., Puentedura, O., Arlener, W., Kastad Hoiskar, B. A., Tellefsen, C. W., Heese, B., Jones, R. L., Aliwell, S. R. & Freshwater, R. A. (1999). Slant column measurements of O<sub>3</sub> and NO<sub>2</sub> during NDSC intercomparison of zenith-sky UV-visible spectrometer in June 1996, *J. Atmos. Chem.*, 32, 281-314
- Sarkissian, A., Roscoe, H. K., & Fish, D. J. (1995). Ozone Measurement by Zenith-Sky Spectrometers: an Evaluation of Errors in Air-Mass Factors calculated by Radiative Transfer Models, *J. Quant. Spectrosc. Radiat. Transfer*, Vol. 54(3), 471 - 480.
- Shettle, E. P., (1989). Models of Aerosols, Clouds and Precipitation for Atmospheric Propagation Studies, *Proceedings of the Electromagnetic Wave Propagation Panel Specialists' Meeting*, pp. 15.1-15.13, ISBN 92-835-0548-4, Copenhagen, Denmark, 9-13 October 1989.
- Solomon, S., Schmeltekopf, A., & Sanders, R. W., (1987). On the interpretation of zenith sky absorption measurements, *J. Geophys. Res.*, Vol. 92, 8311-8319.
- Spiga, G., Boffi, V. C. & Magnavacca, A. (1992). Density Profiles of Charged Test Particles Diffusing in a Slab of Finite Thickness, *Transport Theory and Statistical Physics*, Vol. 21 (4-6), 667-711.
- Stutz, J. & Platt, U., (1997a). Improving long-path differential optical absorption spectroscopy (DOAS) with a quartz-fiber mode-mixer, *Appl. Opt.* Vol. 36, 1105-1115
- Stutz, J. & Platt, U., (1997b). A new generation of DOAS instruments. In: Bösenberg, J., Brassington, D. J., Simon, P. C. (eds.) EUROTRAC Final Report Vol 8: Instrument Development for Atmospheric Research and Monitoring, pp. 370-378
- Takeuchi, K. (1982). Fundamental Theory of the Direct Integration Method for Solving the Steady-state Integral Transport Equation for Radiation Shielding Calculation, *Nucl. Sc. Eng.*, Vol. 80, 536.
- Taylor, J. R. (1982). *An introduction to error analysis : the study of uncertainties in physical measurements*, Oxford university press, Oxford.
- Tomasi, C. & Paccagnella, T. (1986). L'influenza dell'Uomo sul clima del nostro pianeta, II.- Gli effetti delle particelle di aerosol, *Giornale di Fisica*, Vol. XXVII, N.2, (April-June 1986).
- Wagner, T., Ibrahim, O., Sinreich, R., Friess, U., von Glasow, R., & Platt, U., (2007a). Enhanced tropospheric BrO over Antarctic sea ice in mid winter observed by MAX-DOAS on board the research vessel Polarstern, *Atmospheric Chemistry and Physics* Vol.7, 12, 3129-3142.
- Wagner, T., Burrows, J. P., Deutschmann, T., Dix, B., von Friedeburg, C., Frieß, U., Hendrick, F., Heue, K.-P., Irie, H., Iwabuchi, H., Kanaya, Y., Keller, J., McLinden, C. A., Oetjen, H., Palazzi, E., Petritoli, A., Platt, U., Postylyakov, O., Pukite, J., Richter, A.,

- van Roozendael, M., Rozanov, A., Rozanov, V., Sinreich, R., Sanghavi, S., & Wittrock, F., (2007). Comparison of box-air-mass-factors and radiances for Multiple-Axis Differential Optical Absorption Spectroscopy (MAX-DOAS) geometries calculated from different UV/visible radiative transfer models, *Atmos. Chem. Phys.* 7, 1809-1833.
- Walthall, C. L., Norman, J. M., Wes, J. M., Campbell, G. & Blad, B. L. (1985). Simple Equation to Approximate the Bidirectional Reflectance from Vegetation Canopies and Bare Soil Surfaces, *Appl. Optics* 24:383-387
- Young, A. T. (1981). Rayleigh scattering, *Appl. Optics* Vol. 20, 4, 533-534.

# Monte Carlo Simulation of Pile-up Effect in Gamma Spectroscopy

Ali Asghar Mowlavi<sup>1</sup>, Mario de Denaro<sup>2</sup> and Maria Rosa Fornasier<sup>2</sup>

<sup>1</sup>*Physics Department, Sabzevar Tarbiat Moallem University, Sabzevar*

<sup>2</sup>*Struttura Complessa di Fisica sanitaria, A.O.U. "Ospedali Riuniti" di Trieste, Trieste,*

<sup>1</sup>*Iran*

<sup>2</sup>*Italy*

## 1. Introduction

As it is well known, an ideal spectroscopy amplifier should have a constant amplification for pulses of all amplitudes, without distorting any of them. Unfortunately, some pulse distortion is always present because of electronic noise, gain drift due to temperature, pulse pile-up and limitations on the linearity of the amplifier (Knoll, 2000). Therefore, the final spectrum of the detected particles in the detection system is disturbed by these factors. In this chapter, we have concentrated on pile-up effect calculation over the gamma spectrum in a NaI(Tl) scintillation detector.

The fact that pulses from a radiation detector are randomly spaced in time can lead to interfering effects between pulses, which are more likely to occur as the count rate increases. These effects are generally called pile-up. It is known that in a scintillation detector the highest possible radiation counting rate is controlled by the pulse pile-up characteristics of the detection system. In high true count rate (TCR) spectrometry, the pulse pile-up effect is considerable and its spectrum distortion can only be determined by Monte Carlo simulation. (Fazzini et al., 1995; Gostely, 1996; Tenney, 1984).

## 2. Background

Pile-up phenomenon is well known; it can be divided into two general types, which have somewhat different effects on pulse height measurements. The first is known as peak pile-up: it occurs when two or more pulses are sufficiently close together to be treated as a single pulse by the analysis system. The pile-up of pulses has the effect of removing them from the proper position in the pulse height spectrum, and the area under the full-energy peak in the spectrum will no longer be a true measure of the total number of full-energy events. The second type of pile-up is called tail pile-up and involves the superposition of the tail of a pulse to the next one. The effect of tail pile-up on the measurement is to worsen the energy resolution by adding wings to the shape of the recorded peaks. Since tail pile-up does not change the location of acquired events within the energy spectrum, it will not be considered in this work.

In the statistical analysis of pile-up, it is assumed that any inherent dead time of the detector and the preamplifier is small compared with the pile-up resolution time  $\tau$  of the pulse

processing system. This time  $\tau$  is defined as the minimum time that must separate two events to avoid pile-up. Thus, the events which arrive at the amplifier with Poisson distribution are assumed to pile-up only if they occur within a time spacing less than  $\tau$  (Knoll, 2000). In this chapter we describe the Monte Carlo simulation of the pile-up effect distortion over the gamma spectrum of a NaI(Tl) detector. First, MCNP Monte Carlo code was applied to calculate the pulse height spectrum and the detector efficiency. Then, a custom code was written in FORTRAN language to simulate the distortion in pulse height spectrum due to the pile-up effect for paralyzable and nonparalyzable systems by Monte Carlo method. The results of the simulations were compared with the experimental spectra of a  $^{137}\text{Cs}$  source as well as with the experimental measurement of count rate performance of a gamma camera system (Mowlavi et al., 2006).

### 3. Algorithm of pile-up simulation

In the statistical analysis of pile-up, it is assumed that any inherent dead time of the detector and the preamplifier is small compared with the pile-up resolution time  $\tau$  of the pulse-processing system. This time  $\tau$  is defined as the minimum time that must separate two events to avoid pile-up. Thus, the events which arrive at the amplifier with Poisson distribution are assumed to pile-up only if they occur within a time spacing less than  $\tau$ . True events are assumed to occur at a rate  $n$  and, due to pile-up, the recording system will perceive counts at a lower rate  $m$ , so:

$$\begin{aligned} m &= \frac{n}{1 + n\tau} && \text{Nonparalyzable system} \\ m &= ne^{-n\tau} && \text{Paralyzable system} \end{aligned} \quad (1)$$

In general, the probability for a given count to be formed from the pile-up of  $(x+1)$  true events are (Knoll, 2000):

$$\begin{aligned} P(x) &= \frac{(n\tau)^x e^{-n\tau}}{x!} && \text{Nonparalyzable system} \\ P(x) &= e^{-n\tau} (1 - e^{-n\tau})^x && \text{Paralyzable system} \end{aligned} \quad (2)$$

Both the nonparalyzable and the paralyzable formulation have been used in pile-up simulation. The Monte Carlo calculation includes two steps, as described in the diagram of Figure 1. First, MCNP Monte Carlo code (Briesmeister, 2000) was employed to calculate the pulse height spectrum and the detector efficiency for the geometry configuration of the detection system. The simulated experimental configuration with MCNP does not take into account the actual energy resolution of the system as well as does not consider the pile-up effect. In the second step, in order to obtain a more realistic simulation, we wrote a Monte Carlo code by *FORTRAN PowerStation 4.0* software language. The pulse height spectrum coming from the MCNP was convoluted by a Gaussian-spread function corresponding to the measured energy resolution. Then, to consider the pile-up effect, for paralyzable and nonparalyzable systems, we implemented a simulation based on Monte Carlo method. The following analytical function was applied for the pulse shape with  $t_0$  starting time and  $a_0$  amplitude:

$$F(t) = \begin{cases} 4a_0(1-e^{-\frac{(t-t_0)}{\tau_p}}) e^{-\frac{(t-t_0)}{\tau_p}} & t \geq t_0 \\ 0 & t \leq t_0 \end{cases} \quad (3)$$

where  $\tau_p = 0.230 \mu\text{sec}$  is the time constant of NaI(Tl) crystal.

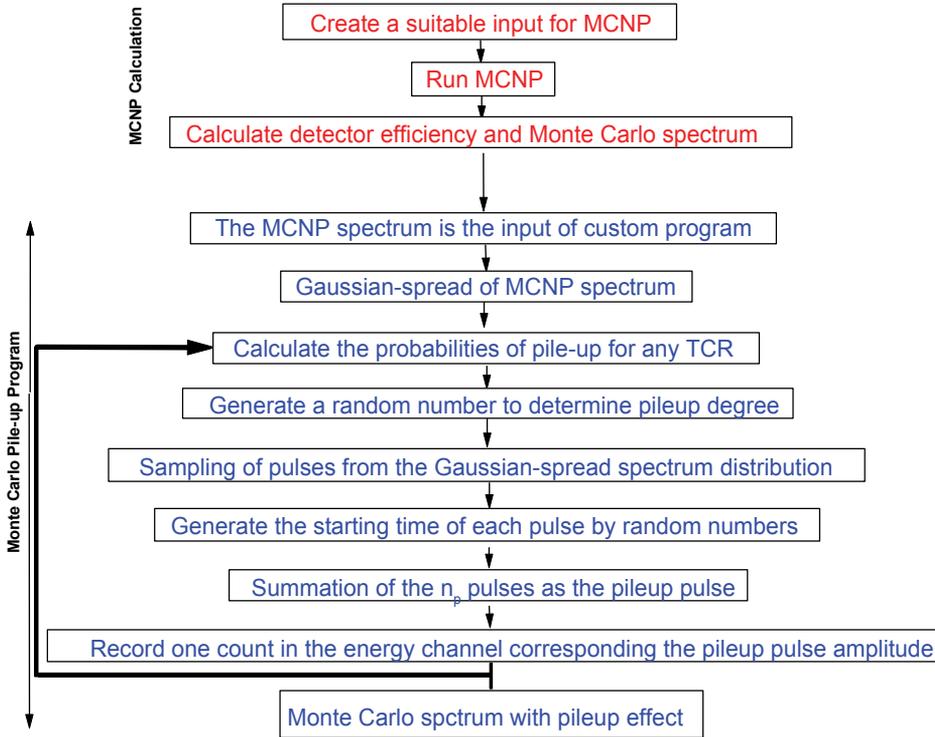


Fig. 1. The diagram of Monte Carlo pile-up calculation algorithm.

The first component of the pulse function,  $(1-e^{-\frac{(t-t_0)}{\tau_p}})$ , corresponds to the increase of the excited states of the NaI(Tl) when the gamma radiation interacts with the crystal, and the second component,  $(e^{-\frac{(t-t_0)}{\tau_p}})$ , shows the decay of the excited states according to the exponential decay law. The summation of three pulses with different amplitude ( $a_0 = 5, 4, 3$ ) and random starting time ( $t_0 = 0.0945, 0.2854, 0.4507$ ) are shown in Figure 2, it can be seen easily that the maximum amplitude of the pile-up pulse is 8.658.

We considered the degree of pile-up as the number of pulses  $n_p$  that made pile-up. For each true count rate (TCR), corresponding to the detector efficiency and source activity, the degree of pile-up is determined, according to the Eq. (2), by a random number ( $r$ ) with uniform distribution in  $[0,1]$  interval. If  $r$  is less than  $P(0)$ , one pulse is sampled from the Gaussian-spread Monte Carlo pulse height spectrum and the event is considered free of pile-up. Otherwise, if:

$$\sum_{i=0}^{k-2} P(i) < r \leq \sum_{i=0}^{k-1} P(i) \quad (4)$$

$n_p$  pulses ( $n_p = k$ ) are sampled from the Gaussian-spread Monte Carlo pulse height spectrum. After the determination of the number of pulses, the starting time of each pulse was simulated by a new random number with uniform distribution in  $[0, \tau]$  interval. Finally the resulting pulse is obtained by linear superposition of the  $n_p$  pulses. We assumed the absolute maximum of the resulting pulse as its amplitude and then we recorded it in the corresponding energy channel.

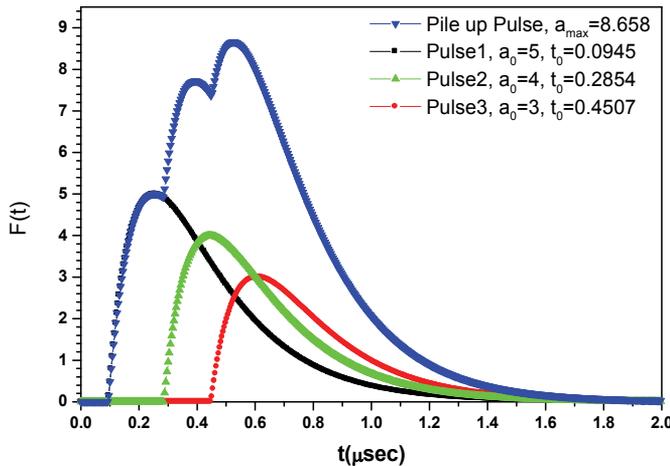


Fig. 2. The summation of the three pulses with different amplitude and random starting time.

#### 4. Calculation of the pile-up distortion on $^{137}\text{Cs}$ pulse height spectrum

##### 4.1 Materials and methods

We have used the pile-up code that we introduced before, to obtain the spectrum with pile-up disturbance and the sub-spectra due to multi pulses pile-up (Mowlavi et al., 2006). Experimental spectrum of  $^{137}\text{Cs}$  source has been measured by a 3in×3in NaI(Tl) scintillation detector in a lead box as shown in Fig. 3. The count rate was about  $1.723 \times 10^5$  per second. MCNP simulated spectrum by F8:p tally and measured spectrum are presented in Fig. 4. Differences of these spectra are clear, especially in the energy region higher than the photoelectric peak.

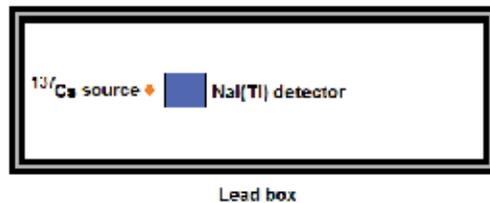


Fig. 3. The geometry set up of the  $^{137}\text{Cs}$  spectrum measurement.

The MCNP pulse height output spectrum is convoluted by a Gaussian function corresponding to the energy resolution of the NaI(Tl) detector.

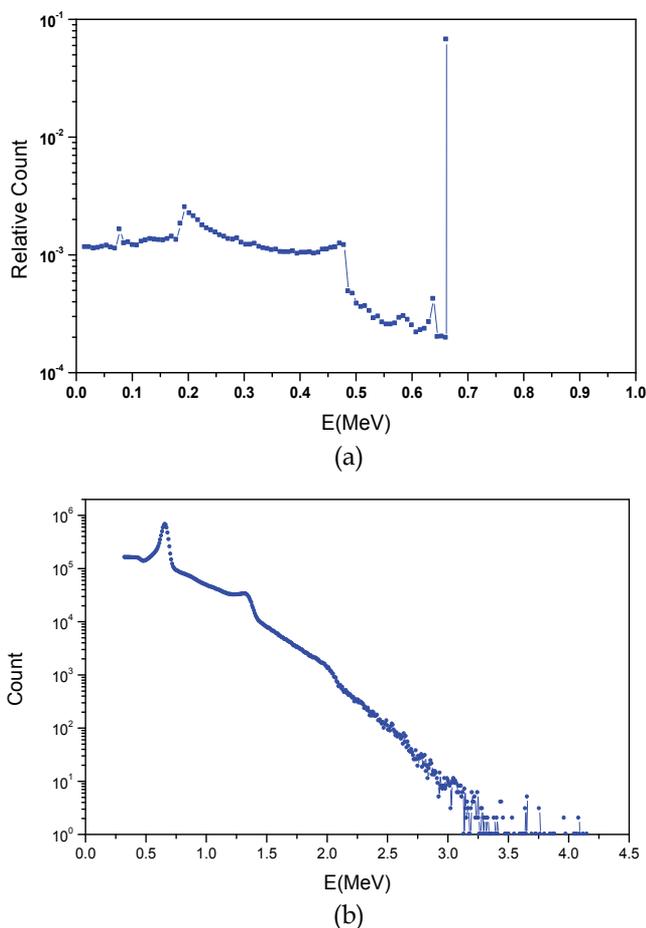


Fig. 4. a) The MCNP pulse height output spectrum; b) the measured spectrum.

#### 4.2 Results and discussion

A comparison of experimental and pile-up computational spectra is presented in Fig. 5. The result shows a good agreement, just a little difference on the left side of photo peak at 0.662 MeV and two peak pile-up at 1.324 MeV which may be caused by tail pile-up. It can be seen the wide peak at 1.986 MeV produced by three peak pile-up. As well as, the full width at half maximum of these peaks are grown by increasing of the pile-up degree. Corresponding to the end point of the measured spectrum the pile-up occurs till sixth degree.

The total Monte Carlo spectrum together with free of pile-up, 2, 3, 4, 5 and 6 pulses pile-up sub-spectra are presented in Fig. 6. It must be mention that these sub-spectra can be obtained just by Monte Carlo simulation. About 74.77 % count in the photo peak is due to free pile-up events, 23.27% is due to two pulses pile-up, 1.88% comes from three pulses pile-up, and other portions are neglectable. It means that the main impurity under the photoelectric peak is due to two pulses pile-up.

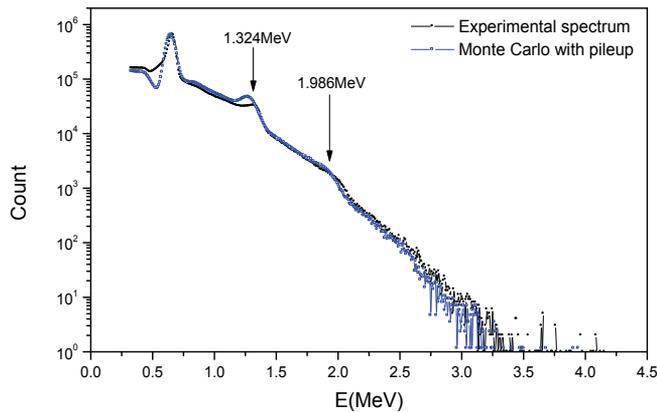


Fig. 5. Comparison of the experimental and pile-up computational spectra.

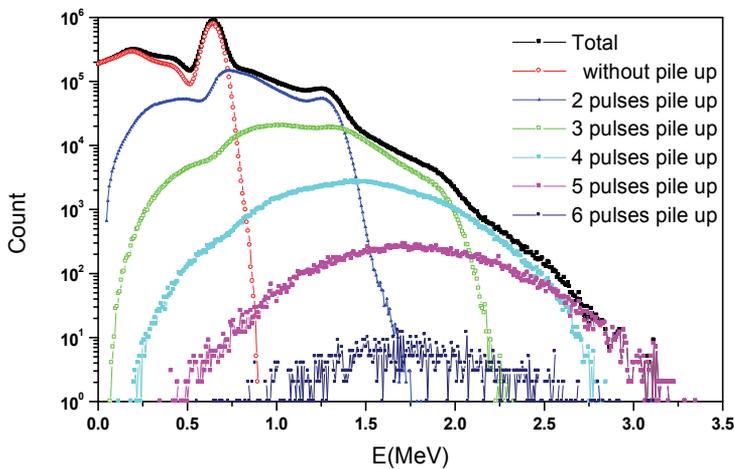


Fig. 6. The total MC spectrum together with the free of pile-up, 2, 3, 4, 5 and 6 pulses pile-up sub-spectra.

## 5. Monte Carlo simulation of intrinsic count rate performance of a scintillation camera

### 5.1 Introduction

Scintillation gamma camera is a medical equipment, therefore it must be subjected to periodical reference tests. Intrinsic count rate is considered an important parameter to be measured during the acceptance test of the instrumentation and for the periodical monitoring of the system.

The National Electrical Manufacturers Association (NEMA) and the International Atomic Energy Agency (IAEA) have published protocols for Quality Controls of Nuclear Medicine instrumentations with a detailed description of the experimental methods for the measurement of the intrinsic count rate performance (NEMA, 1994; IAEA, 1984). Based on the detailed description of the measurement procedures, it is possible to simulate the experimental condition by a Monte Carlo code analysis.

In the present work, MCNP code is used to calculate the pulse height spectrum and the gamma efficiency of a single crystal gamma camera. The simulated data so obtained are then processed by custom software to take into account the pile-up effect, a phenomenon which, at high count rate, produces a distortion in the detector spectrum which cannot be neglected.

To evaluate the consistency of the simulation, we compared the experimental measurement of intrinsic count rate performance with the Monte Carlo result.

## 5.2 Experimental measurement of intrinsic count rate

Intrinsic count rate test is an important measurement to evaluate the performance of a scintillation camera in terms of its response to an increasing incident gamma radiation.

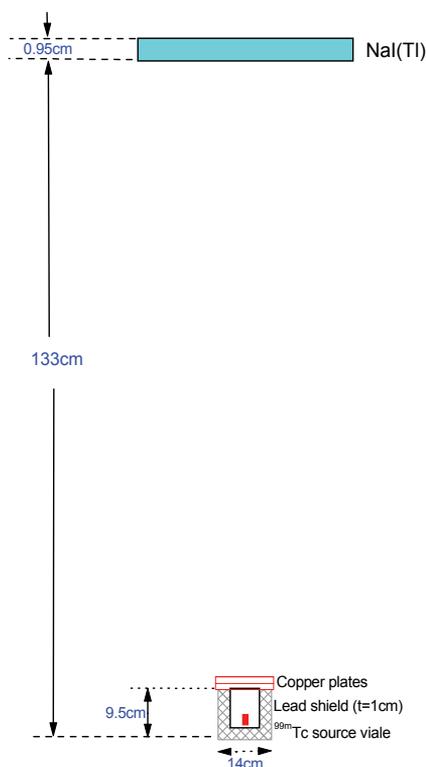


Fig. 7. The configuration of count rate test of gamma camera.

Our measurement was carried out on a gamma camera Siemens E. Cam equipped with a 9.5 mm thick NaI(Tl) single crystal detector. Following the mentioned protocols (NEMA, 1994; IAEA, 1984), we employed a  $^{99m}\text{Tc}$  source contained in a small vial (about 10MBq) with 22 copper absorbers, each 2 mm thick, placed over the source. The gamma camera was set with a 20% window at the 140.5 keV peak. Window Observed Count Rates (WOCR) was collected for 1 minute acquisitions, starting with all the absorbers in place over the source and then removing the uppermost absorbers one by one. This procedure increases the incident gamma radiation flux and the input count rate in inverse proportion to the attenuation factor of the absorber removed. Usually calibration of absorbers is calculated by

measuring, in low count rate condition, the ratio between counts with the plate over the source and without it. In case of uniformity of the thickness of the copper plates, a mean attenuation factor  $\bar{A}$  can be used (NEMA, 1994).

For a number  $k$  of copper attenuation plates in place:

$$N(0) = A_1^{-1} A_2^{-1} A_3^{-1} \dots A_k^{-1} R(k) = R(k) \prod_{i=1}^k A_i^{-1} \quad (3)$$

where:

$N(0)$  = window true count rate(WTCR) calculated for  $k=0$  absorber plates

$A_i$  = individual attenuation factor of the copper plate number  $i$

$R(k)$  = window observed count rate(WOCR) for  $k$  plates

The configuration of the count rate test of a gamma camera has been shown in Fig. 7. Fig. 8 shows the experimental result of the count rate performance. In our measurement we evaluate the slope of the straight line of  $\ln(\text{WOCR})$  against  $k$  (number of plates over the source) in the low count rate range to determine the mean attenuation factor,  $\bar{A}$  (Geldenhuys et al., 1988).

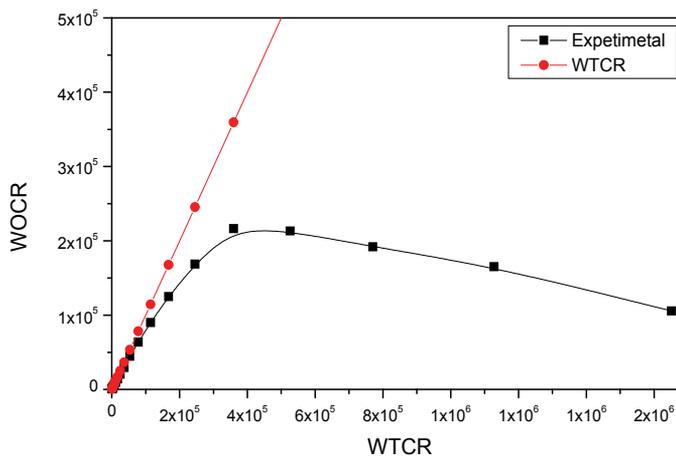


Fig. 8. The experimental result of count rate performance.

### 5.3 Monte Carlo simulation of intrinsic count rate performance

The Monte Carlo calculation includes two steps as describe in the diagram of Fig. 1.

First, MCNP Monte Carlo code (Briesmeister, 2000) was employed to calculate the pulse height spectrum and the detector efficiency in detecting 140.5 keV gamma rays, for a NaI(Tl) crystal and the geometrical configuration shown in Fig. 7, for various numbers of copper plates removed.

The three MCNP calculated spectra corresponding 1, 5, and 10 copper plates over the source are shown in Fig. 10-a, also Fig. 10-b shows the Gaussian spread of these spectra. It is clear that the gamma spectrum is varying with the number of copper plates over the source. Fig. 11 shows the total count per on particle from the source against the number of copper plates over the source.

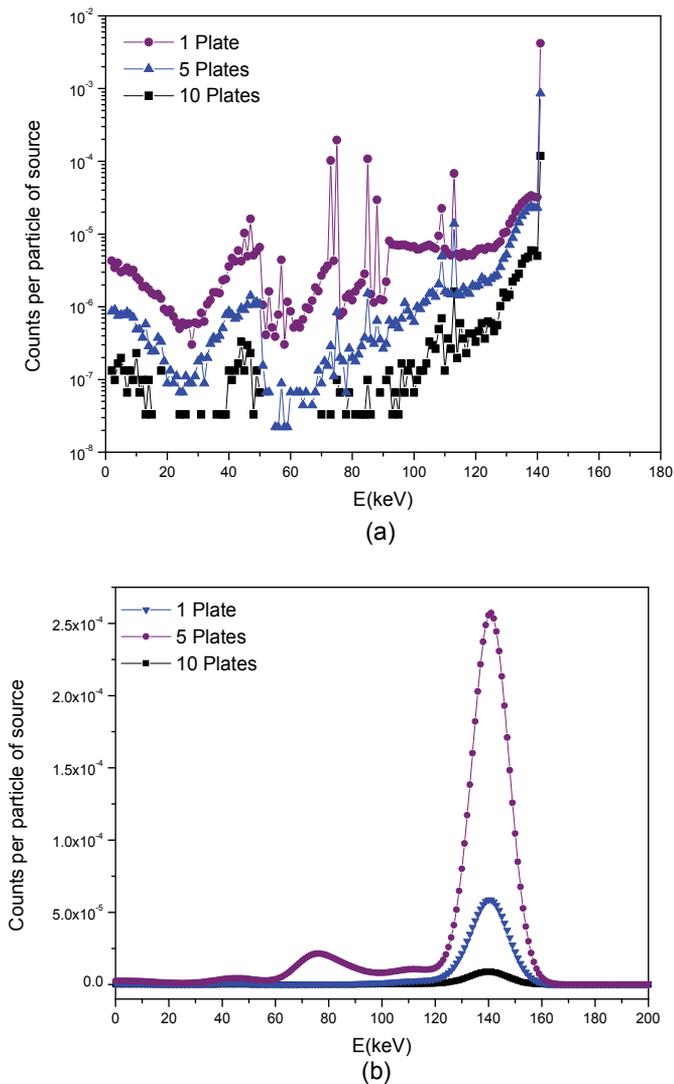


Fig. 10. a) The three MCNP calculated spectra corresponding to 1, 5, and 10 copper plates over the source, b) the Gaussian spread of the spectra.

The simulated experimental configuration with MCNP does not take into account the actual energy resolution of the system as well as does not consider the pile-up effect.

In the second step, in order to obtain a more realistic simulation, we used the Monte Carlo pile-up code. First the pulse height spectrum coming from the MCNP was convoluted by a Gaussian-spread function corresponding to the measured 11.2% gamma camera energy resolution. Fig. 12 shows the experimental spectrum of the gamma camera crystal with 11.2% energy resolution at 140.5 keV peak. Then, to consider the pile-up effect, for paralyzable and nonparalyzable systems, we implemented a simulation based on Monte Carlo method. In the next section, Monte Carlo simulation has been done for  $\tau = 0.5, 1, 1.5$   $\mu$ sec values.

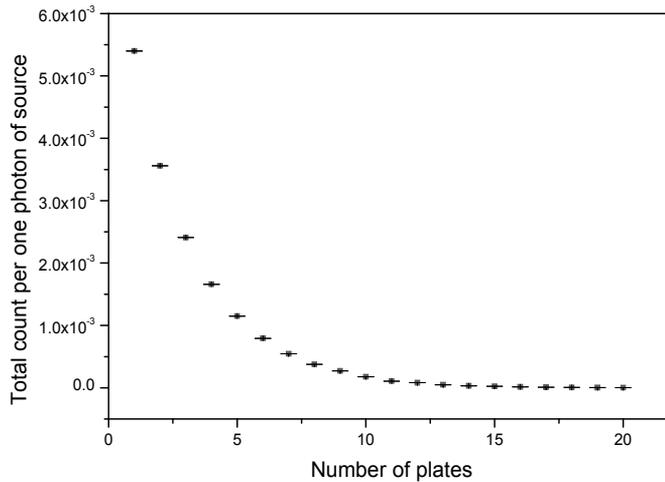


Fig. 11. The total count per one particle from the source against the number of copper plates over the source.

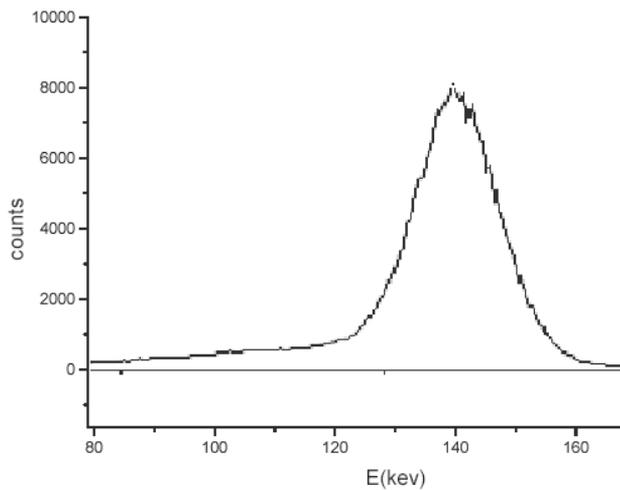


Fig. 12. The experimental spectrum of the gamma camera crystal at low count rate for a  $^{99m}\text{Tc}$  source.

#### 5.4 Results and discussion

We recorded the MCNP result of the detector efficiency ( $R$ ) in the second column of Table 1. The third column shows the activity of  $^{99m}\text{Tc}$  source ( $A_n$ ) in the count rate test measurement. We have calculated the pulse height spectrum without pile-up in very low count rate  $\text{TCR}=1$ , and the pulse height spectrum with pile-up in  $\text{TCR}=R \times A_n$ , for any number of plates. The window observed count rate is the counts in the 20% window considering the pulse height spectrum with pile-up and the window true count rate is the counts in the 20% window considering the pulse height spectrum without pile-up (Table 1, columns 5, 6 and 7).

No. of Plates	R (%)	$A_p$ (Bq)	TCR (k count)	WTCR (k count)	WOCR Paralyzable (k count)	WOCR Nonparalyzable (k count)
1	0.540±0.001	3,65E8	1969.56±1.40	1581.94±1.26	77.26±0.28	178.43±0.42
2	0.357±0.001	3,69E8	1315.26±1.14	1124.16±1.06	162.17±0.40	216.88±0.47
3	0.241±0.001	3,69E8	890.32±0.94	777.46±0.88	194.36±0.44	227.23±0.48
4	0.166±0.001	3,70E8	613.12±0.78	541.34±0.74	209.24±0.46	220.92±0.47
5	0.115±0.001	3,71E8	425.52±0.65	371.88±0.61	200.71±0.45	209.83±0.46
6	0.079±0.001	3,72E8	291.77±0.54	259.73±0.51	160.99±0.40	172.60±0.42
7	0.055±0.001	3,73E8	203.21±0.45	178.81±0.42	125.65±0.34	127.51±0.36
8	0.037±0.001	3,74E8	139.74±0.34	123.98±0.35	96.64±0.31	97.38±0.31
9	0.027±0.001	3,75E8	99.84±0.32	87.20±0.30	76.49±0.28	76.96±0.28
10	0.018±0.001	3,75E8	66.76±0.26	56.96±0.24	50.75±0.23	50.83±0.23
12	0.0082±0.0001	3,78E8	31.16±0.18	27.37±0.17	25.30±0.16	25.46±0.16
15	0.0026±0.0001	3,81E8	10.02±0.10	8.37±0.09	8.00±0.09	8.10±0.09
20	0.0004±0.0001	3,86E8	1.47±0.04	1.24±0.03	1.13±0.03	1.14±0.03

Table 1. Simulation results for count rate performance.

Fig. 13 shows the Monte Carlo spectrums without pile-up effect and with pile-up effect in paralyzable and nonparalyzable detection systems for 3 copper plates. We like to note that in high TCR, the rate of pulses that escape from the energy window due to the pile-up is higher than the WOCR. For example, in paralyzable system, with tree copper plates, the number of pulses per second which, due to the pile-up escape from the energy window is 3.07 times WOCR. Also, the contribution in WOCR from the pile-up coming from pulses with energies lower than the window interval is negligible. In nonparalyzable system, in the same condition, the number of pulses per second that escape from the energy window due to the pile-up is 2.46 times WOCR and the count rate in the energy window due to the pile-up of low energy pulses is negligible. So, in high TCR, for both paralyzable and nonparalyzable systems, the effect of pulse escaping from the window is more important than the increase of background due to the low energy pulse pile-up.

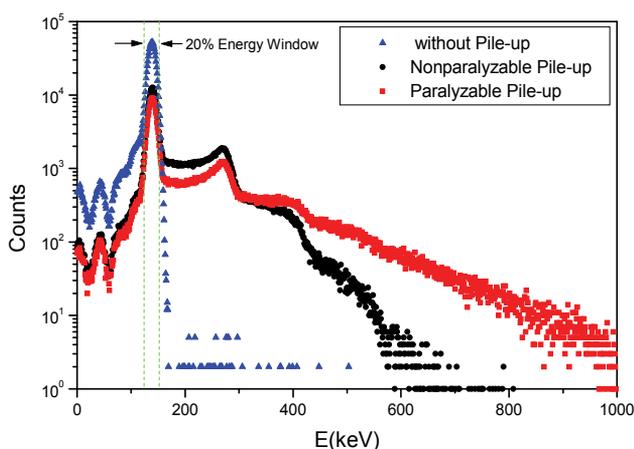


Fig. 13. The MC pulse height spectrum for three copper plates without and with pile-up in both modes.

In Fig. 14 it is reported the comparison of the pile-up defect in paralyzable and nonparalyzable systems and the sub-spectra of increasing number of pulse pile-up. In order to find a good selection value for  $\tau$ , we have done the count rate performance simulation for three values:  $\tau=0.5, 1, 1.5 \mu\text{sec}$ . Fig. 15 shows the Monte Carlo and the experimental result of count rate performance.

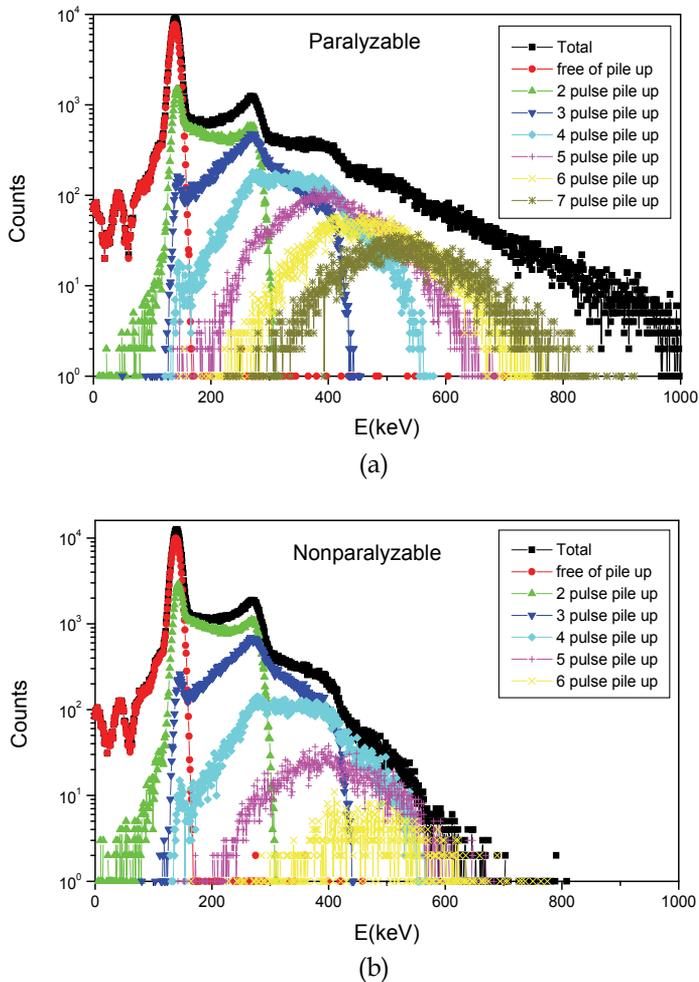


Fig. 14. Monte Carlo spectra with pile-up effect for 3 copper plates, (a) paralyzable and (b) nonparalyzable systems. Sub-spectra of subsequent pile-up pulses are also included.

The comparison of Monte Carlo and experimental results guides us to conclude that the Monte Carlo result for  $\tau$  around  $1 \mu\text{sec}$  is in good agreement with the experimental result for paralyzable mode. In fact, we have calculated the results for  $\tau=0.70, 0.75, 0.80 \mu\text{sec}$  in paralyzable mode as shown in Fig. 16; and the best agreement obtained is for  $\tau=0.75 \mu\text{sec}$  (see Fig. 17). If we consider the nonparalyzable mode for the detection system, the results show a not bad agreement for  $\tau=2.5 \mu\text{sec}$ , although around the peak of WOCT the difference between experimental and computational data is considerable in this mode.

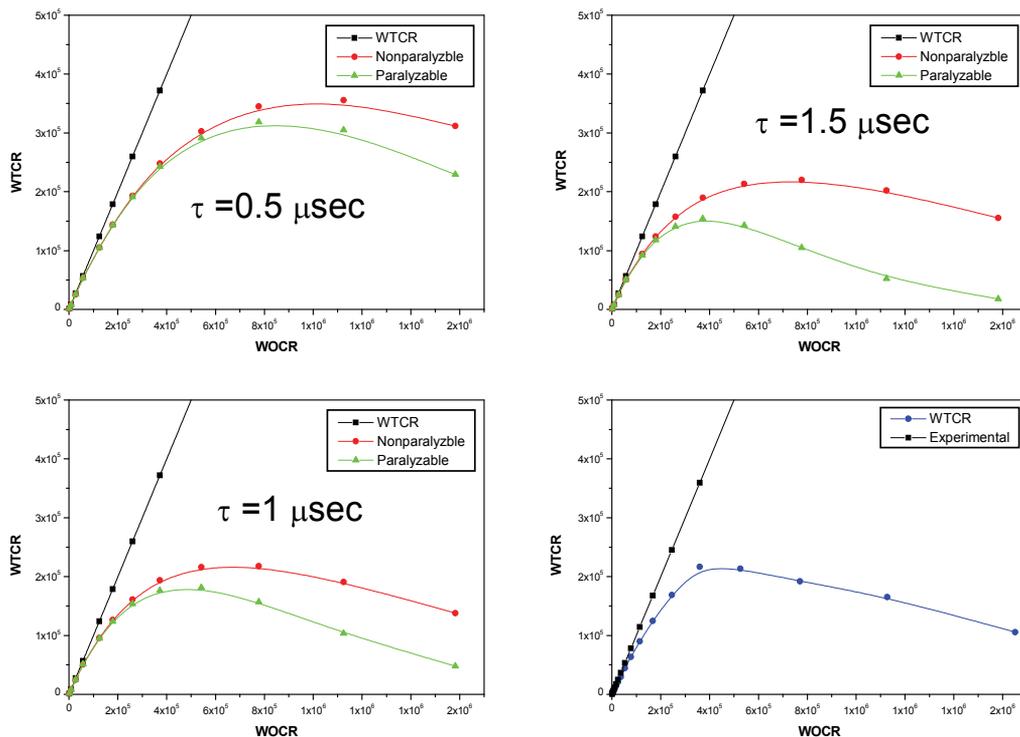


Fig. 15. The experimental and Monte Carlo result for count rate performance of nonparalyzable and paralyzable systems (all of results are shown in the same scale).

Finally, we must mention that the method can also be employed for other detectors and other counting systems (Sjöland et al., 1999; Wu et al., 1996), because it is easily customizable simply substituting some input parameters.

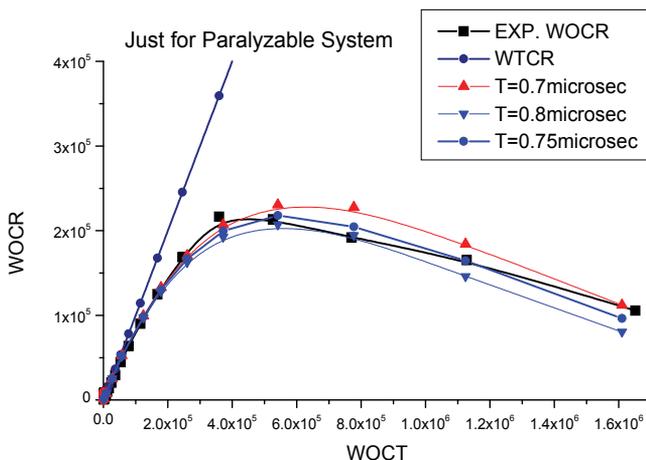


Fig. 16. The experimental and Monte Carlo result for paralyzable mode and  $\tau=0.70, 0.75, 0.80 \mu\text{sec}$ .

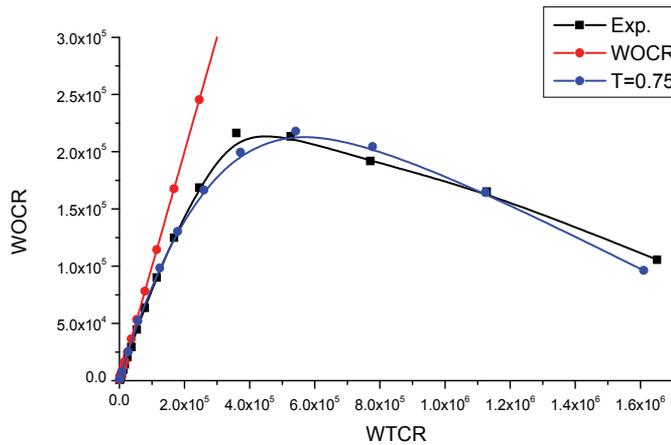


Fig. 17. The best agreement between the experimental and Monte Carlo result for paralyzable mode and  $\tau=0.75 \mu\text{sec}$ .

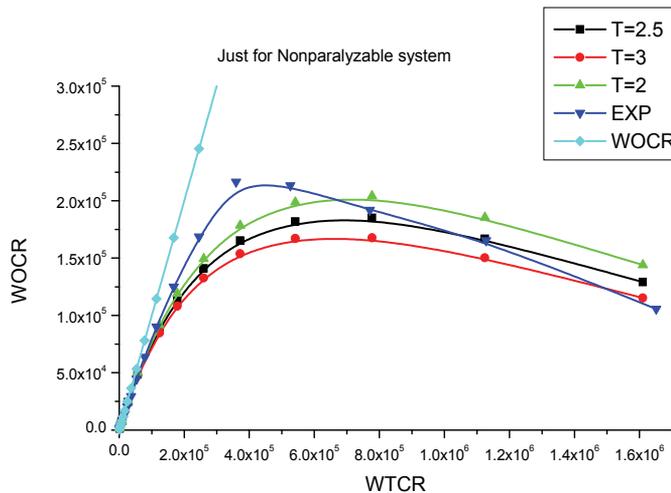


Fig. 18. The experimental and Monte Carlo result for nonparalyzable mode and  $\tau=2, 2.5, 3 \mu\text{sec}$ .

## 6. Conclusion

The Monte Carlo pile-up code that we developed can be used to correct pile-up distortion in gamma spectroscopy.

At the first, we have compared and analyzed the experimental and computational spectrum considering pile-up effect of a  $^{137}\text{Cs}$  source in a NaI(Tl) detector. The interesting sub-spectra have been obtained, which construct the MC spectrum with pile-up disturbance. The results have shown that the main background in the photoelectric peak is due to two pulses pile-up, about 23.27% of counts. In high count rate situation the total count of the photoelectric peak reduces effectively due to removal of pulses by pile-up.

The second application of the pile-up code was used for gamma camera count rate performance test. The Monte Carlo simulation demonstrated that the gamma camera we tested behaves in presence of a high true count rate as a paralyzable system rather than as a nonparalyzable system. In low true count rate, paralyzable and nonparalyzable simulation results are close to each other. The result obtained by simulation of pile-up with this approach showed a good agreement with experimental measurement. As a future step, we plan to apply the code on the gamma spectrum of a Prompt Gamma Neutron Activation Analysis (PGNAA) system.

## 7. Acknowledgement

The authors would like to thank Prof. G. Furlan and D. Treleani in TRIL program at ICTP, Trieste, Italy, for their contribution in this work.

## 8. References

- AIEA (1984). Quality Control of Nuclear Medicine Instruments Protocol, IAEA, Vienna.
- Briesmeister, J.F. (editor), (2000). MCNP<sup>TM</sup>- A general Monte Carlo N-particle transport code, Version 4C, Los Alamos National laboratory Report LA-13709-M.
- Fazzini, T., Poggi, G., Sona, P., Taccetti, N. (1995). Detailed simulation of pile-up effects in single spectra acquisition. Nucl. Instrum. Meth. Phys. Res. A 356, 319.
- Geldenhuys, E.M., Lotter, M.G., Minnaar, P.C. (1988). A new approach to NEMA scintillation camera count rate curve determination. J. Nucl. Med., 29(4), 538.
- Gostely, J.J. (1996). Pile-up effect for a counting channel with amplitude selection and imposed non-cumulative dead time. Nucl. Instrum. Meth. Phys. Res. A 369, 397.
- Guo, W., Lee, S.H., Gardner, R.P., 2004. The Monte Carlo approach MCPUT for correcting pile-up distorted pulse-height spectra. Nucl. Instr. and Meth. A 531(3), 520-529.
- Knoll, G.F. (2000). Radiation Detection and Measurement, 3rd ed. Wiley, New York.
- Möllendorff, U. V., Giese, H. (2003). Experimental tests of dead-time corrections. Nucl. Instrum. Meth. Phys. Res. A 498, 453.
- Mowlavi, A.A., de Denaro, M., Fornasier, M.R., Binesh, A., 2006. Monte Carlo simulation of intrinsic count rate performance of a scintillation camera for diagnostic images. Appl. Radiat. Isot. 64(3), 390-395.
- Mowlavi, A.A., Koochi-Fayegh, R., 2005. Tally modifying of MCNP and post processing of pile-up simulation with time convolution method in PGNAA. Nucl. Instr. And Meth. A 552(3), 559-565.
- NEMA (1994). Performance Measurements of Scintillation Cameras, National Electrical Manufactures Association (NEMA), Washington, Standards Publication NU 1.
- Palomba <sup>a</sup>, M., D'Erasmo, G., Pantaleo, A. (2003). An application of the CSSE code: analysis of pulse pile-up in NaI(Tl) detectors used for TNA. Nucl. Instrum. Meth. Phys. Res. A 498, 397.
- Palomba <sup>b</sup>, M., D'Erasmo, G., Pantaleo, A. (2003). The Monte Carlo code CSSE for the simulation of realistic thermal neutron sensor devices for Humanitarian Demining. Nucl. Instrum. Meth. Phys. Res. A 498, 384.
- Sjöland, K.A., Munnik, F., Chaves, C., Wätjen, U. (1999). Time-resolved pile-up compensation in PIXE analysis with list-mode collected data. Nucl. Instrum. Meth. Phys. Res. B 150, 69.

- Tenney, F. H. (1984). Idealized pulse pile-up effects on energy spectra. *Nucl. Instrum. Meth. Phys. Res.* 219, 165.
- Woldeselassie, T. (1999). Modeling of scintillation camera systems. *Med. Phys.* 26(7), 1375.
- Wu, X., Brown, J.K., Kalki, K, Hasegawa, B.H. (1996). Characterization and correction of pulse pile-up in simultaneous emission-transmission computed tomography. *Med. Phys.* 23(4), 569.

# Monte Carlo Simulations of Microchannel Plate–Based, Time-Gated X-ray Imagers

Craig A. Kruschwitz and Ming Wu  
*National Security Technologies, LLC*  
USA

## 1. Introduction

In the last two decades, high-speed, time-gated Microchannel plate (MCP) x-ray detectors have proven to be powerful diagnostic tools for two-dimensional, time-resolved imaging and time-resolved x-ray spectroscopy in the field of laser-driven inertial confinement fusion and fast Z-pinch experiments (McCarville et al., 2005; Oertel et al., 2006; Bailey et al., 2004). These detectors' quantitative measurements are critical for a comprehensive understanding of the experimental results. To assist their characterizations and to aid design improvements, a more comprehensive Monte Carlo simulation model for the MCP detector is needed.

The MCP detectors are widely used as electron, ion, and x-ray detectors, as well as imaging tools in many areas of scientific research. Their principles of operation have been documented in the literature (Wiza, 1979; Fraser et al., 1982; Fraser et al., 1984; Kilkenny, 1991) as have extensive research efforts to characterize detection sensitivity (Ze et al., 1999; Landen et al., 1993; Landen et al., 1994), angular and energy dependences (Hirata et al., 1992; Landen et al., 2001; Rochau et al., 2006), and temporal and spatial resolution (Robey et al., 1997). In many previous studies, a discrete dynode gain model was used to describe the MCP gain dependence on the applied voltage (Eberhardt, 1979). This dependence is extremely nonlinear. The discrete dynode model assumes that the MCP can be treated as a conventional, discrete-stage electron multiplier with a fixed number of stages. This gain model uses a few free parameters, chosen to best fit a certain MCP's data. The discrete dynode model seems to work well to describe the behavior of MCPs under some circumstances, but several factors are not included when inferring the secondary electron yield from gain measurements. For example, the discrete dynode model assumes that the number of dynode stages is independent of the applied voltage on the MCP (the number of stages is chosen to best fit the gain vs. voltage data), which is unlikely to be valid.

In addition to the discrete dynode model, previous researchers have also performed Monte Carlo-based computer simulations of the MCP response to a steady-state voltage for straight and tilted microchannels (Guest, 1971; Ito et al., 1984; Choi & Kim, 2000; Price & Fraser, 2001). These simulations apparently did not include the constraint of energy conservation for the secondary electrons. This constraint prevents the aggregate energy of the emitted electrons from exceeding the primary electron energy. Furthermore, these previous efforts omit the effects of low-energy electrons' elastic scattering from the channel walls, potentially an important effect for the low-impact energies prevalent in an MCP electron cascade. A further difficulty encountered by all such previous efforts (and, indeed,

the present one) is that a fair number of adjustable parameters in these simulations cannot be unambiguously determined from the existing data. Consequently, inconsistent parameter sets appear in the literature, with some more robustly constrained by the data than others. This chapter is divided into several parts. To begin we present a somewhat detailed description of our simulation model, which builds upon our earlier work (Wu et al., 2008; Kruschwitz et al., 2008). Section 2 lays out the secondary emission equations used in the simulations and briefly describes how we approximate MCP saturation. Results from various sets of simulations of MCPs under both steady state and pulsed bias voltages are shown. Comparisons of simulated and measured MCP gain in DC mode appear in Section 3. MCP performance under an ideal square waveform pulse is described in Section 4. An example of how to use Monte Carlo simulation to optimize MCP detector design is shown in Section 5, and is followed by a detailed comparison of the simulations with experimental results in pulsed mode to better evaluate the effectiveness of the model. Comparisons between experimental and simulation results of detector optical gate profile and sensitivity uniformity, as well as detector gain under pulsed operation are discussed in Sections 6 and 7. MCP saturation in DC and pulsed modes is addressed in Section 8. Spatial resolution of the MCP detector is discussed in Section 9. We conclude with some simulations of small-pore ( $2\ \mu\text{m}$ ) MCPs, making some predictions about their potential performance in x-ray imaging systems in Section 10

## 2. Model and computational algorithm

An MCP is essentially an array of parallel continuous electron multipliers. A schematic of an MCP x-ray detector is shown in Fig. 1.

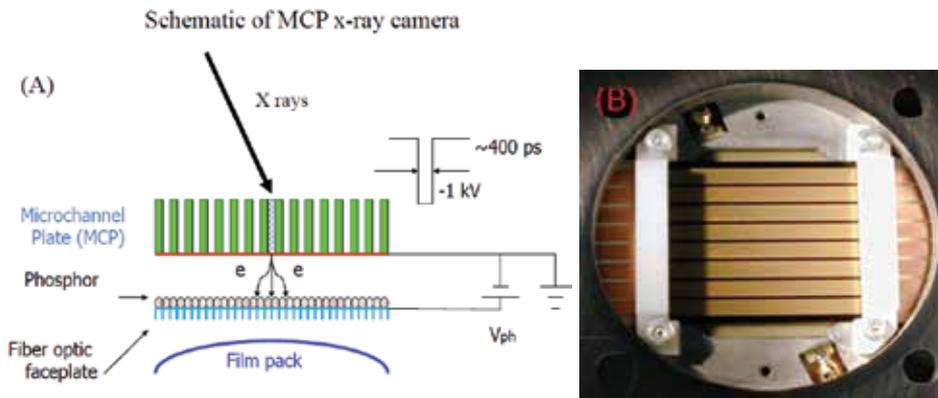


Fig. 1. (A) Schematic of a high speed time-gated MCP based x-ray detector. The MCP is usually biased by a negative voltage while the phosphor is biased by a positive voltage. (B) Eight-strip MCP detector (NSTec H-CA-65).

For our work the individual MCP channels were typically  $\sim 10\ \mu\text{m}$  in diameter, with a  $12\ \mu\text{m}$  spacing between pore centers (denoted  $D_{cc}$  throughout), and with a plate thickness of  $450\text{--}600\ \mu\text{m}$ ; recently, small-pore MCPs with pore diameters as small as  $2\ \mu\text{m}$  ( $D_{cc} = 3\ \mu\text{m}$ ) have become available, and these are studied here as well. Generally, the channels are set at an angle,  $\alpha$ , relative to the MCP surface normal, called the bias angle. The MCP is also characterized by the ratio of its thickness to the channel diameter, or its  $L/D$  ratio.

We assume that MCP electron cascade dynamics can be approximated by the behavior of a single microchannel. For our purposes, it is acceptable to assume that all microchannels in a particular plate are identical because we are neglecting any cross-talk effects between adjacent channels. Note, however, that previous researchers have reported that when the MCP is operated at high count rates or used for hard x-ray detection, effects between neighbouring channels may occur (Fraser et al., 1993; Shikhaliyev, 1997). Also, pore-bleaching effects, where the fields generated in a pore affect the gain in adjacent pores, are not used in these simulations; they are assumed to be negligible in subnanosecond MCP gating for the detection of soft x-rays, which is our focus.

The cascade is taken to be initiated by some number of incident electrons. The actual number of primary electrons is sampled from a Poisson distribution. The primary electrons are assumed to be generated by interactions of x rays or UV photons with the reduced lead glass channel surface. The directions in which the primary electrons are emitted are given by a pair of angles for each electron. The first of these angles is defined relative to the surface normal and is assumed to obey a cosine probability distribution:

$$P(\theta) \propto \cos(\theta) \quad (1)$$

where  $\theta$  is the angle relative to the surface normal. The second angle is an azimuthal angle sampled from a uniform distribution between 0 and  $2\pi$ . The initial energies of the primary electrons are somewhat more difficult to determine and are expected to be different for x-ray and UV sources. For x rays, we presume the energies ( $E_s$ ) are sampled from the following probability distribution (Scholtz et al., 1996):

$$f(E_s) = C \cdot \exp\left\{-\frac{[\ln(E_s / E_0)]^2}{2\sigma^2}\right\} \quad (2)$$

where  $E_0 = 2.3$  eV, the most probable energy,  $\sigma = 0.65$ , and  $C$  is a normalization constant. The same probability distribution is used to determine the initial energies of the secondary electrons. We use this distribution in lieu of measurements of the energy distribution of electrons produced from the interaction of keV x rays with the MCP glass. The values for  $E_0$  and  $\sigma$  approximately match the experimental data of Authinarayanan & Dudding (1976) on MCP glass secondary electron emission.

For UV photons, the initial energies of the primary electrons are determined by the work function of the reduced lead glass material of the MCP and the UV photon energy, the value of the photon energy being 6.2 eV in our experiments. The work function of the reduced lead glass is poorly known, but values of around 5 eV have been quoted in the literature (Melamid, 1972). Therefore, we deduce that photoelectrons produced by the UV photons have an initial energy of  $\sim 0.1$ – $1.2$  eV. Simulation results with primary electron energies in this range match our data quite well.

In this model, we assume that only the photoelectrons and Auger electrons generated adjacent to the microchannel can initiate a cascade within the channel, in agreement with experimental observation for  $<5$  keV x rays (Rochau et al., 2006). The initial photoelectrons are generated in a timeframe ( $<10^{-12}$  s) that is much shorter than the transit time of the cascading electrons. The time lag of secondary emission is estimated to be  $10^{-13}$  to  $10^{-14}$  s, much shorter than any timescale relevant to electron-cloud effects. Thus, we suppose that

secondary electrons are generated instantaneously when a primary electron hits the lead glass surface. For the simulations presented in this chapter, all electron cascades begin at the MCP input face. For each electron incident on the channel wall, equations dependent on the incident energy and angle are used to determine the mean secondary emission yield.

The actual value of the average secondary emission yield for a primary electron incident on the MCP channel surface at angle  $\theta_i$  and with energy  $V_i$  is determined by (Vaughan, 1989)

$$\delta(V_i, \theta_i) = \delta_m(\theta_i) \left( \frac{V_i}{V_m(\theta_i)} \exp \left[ 1 - \frac{V_i}{V_m(\theta_i)} \right] \right)^s \quad (3)$$

where  $s$  is a free parameter whose value is chosen to best fit the data (we use  $s = 0.62$ ), and  $V_m(\theta_i)$  and  $\delta_m(\theta_i)$  are given by the following equations (Vaughan, 1989):

$$V_m(\theta_i) = V_m(0) \left( 1 + k \frac{\theta^2}{2\pi} \right) \quad (4)$$

$$\delta_m(\theta_i) = \delta_m(0) \left( 1 + k \frac{\theta^2}{2\pi} \right) \quad (5)$$

where  $V_m(0)$  is the impact energy yielding the maximum mean secondary electron yield  $\delta_m(0)$ , at normal incidence, and  $k$ , a constant determined by the data (between 0 and 2), is usually a surface smoothness indicator. Values between 0.5 and 1 fit our data well. We take  $V_m(0)$  to be  $\sim 300$  eV and  $\delta_m(0)$  to be 3–4, as in the experimental results of Authinarayanan & Dudding (1976). Final values are chosen to match our measurements of the MCP sensitivity versus static bias voltage. The actual secondary yield is determined by random sampling of a Poisson distribution with a mean value of  $\delta(V_i, \theta_i)$ , determined using Eq. (3).

Other authors (Guest, 1971; Ito et al., 1984; Choi & Kim, 2000; Price & Fraser, 2001) have used a variety of alternate forms for Eqs. (3)–(5), each with their own adjustable parameters. The manners in which these models simulate the secondary emission processes in an MCP are very similar. This is largely because the majority of electron-channel wall collisions are low energy ( $< 50$  eV) a domain in which each of these models is nearly linear and for which there is little measured data. Furthermore, measurements of the dependence of secondary emission on the incident angle of the primary are typically at angles between zero degrees (normal incidence) and 60–70 degrees. In the channel, however, even for low bias voltages, most primaries impact at a near grazing angle ( $\sim 70$ –80 degrees), where there is essentially no data. Thus, the published models are little more than useful starting points for modelling secondary emission in the MCP.

Knowing the number of secondary electrons, each electron's initial energy is assigned by sampling Eq. (2). Note that various forms for the secondary energy distribution have been used by different authors (Guest, 1971; Ito et al., 1984; Choi & Kim, 2000; Price & Fraser, 2001). We have found that the specific form of the distribution generally has little effect on the simulation results. Of greatest importance to the outcome is the value of the most probable energy. While we assume the emission energies of the secondary electrons to be uncorrelated, we do require that energy is conserved by ensuring that the sum of the secondary electron emission energies be less than the impact energy of the primary electron. As it happens, a straightforward sampling of Eq. (2) only very rarely violates the

conservation condition, thus we implement energy conservation by repeatedly sampling the secondary electron energies from Eq. (2) until the conservation condition is met.

The direction in which each secondary electron is emitted is sampled from a cosine distribution (Eq. (1)). We assume that the secondary electrons' emission angles are fully uncorrelated, independent of the incident energy and angle, and uncorrelated with the emission energies. Experiments show these to be reasonable conclusions (Bruining, 1954). The trajectories of the secondary electrons are then calculated using nonrelativistic equations of motion. Because the maximum electron kinetic energies are  $\leq 1$  keV, the problem can safely be treated nonrelativistically. After the electron equations of motion are solved, the electron's impact energy and angle with the channel wall are determined, giving the initial conditions for the next generation of electrons. This process repeats until the electrons emerge from the output channel end or the cascade dies out, yielding no output electrons. The positions and velocities at the output of the channel are determined for those electrons that emerge from the channel. Finally, these electrons are then accelerated by the voltage,  $V_{ph}$ , applied between the MCP output face and the phosphor to determine their final positions at the phosphor. In our model electron scattering from the phosphor has not been taken into account.

In previous Monte Carlo simulations (Guest, 1971; Ito et al., 1984; Choi & Kim, 2000; Price & Fraser, 2001), the possibility of the incoming electrons' elastic reflection was neglected. The effect was first considered in our previous work (Wu et al., 2008), and we continue to include the effect. In our simulations elastic scattering was an important effect, particularly for low-bias voltages. It was necessary to include elastic scattering in order to use one parameter set to describe the MCP gain variation with bias voltage over the full range of bias voltages of interest. The probability of electron reflections from lead glass at normal incidence as a function of energy was studied by Scholtz et al. (1996). They discovered that for 10 eV primary electrons incident normal to a lead glass surface, the majority of the resulting secondary electrons were elastically reflected primary electrons, decreasing to just a few percent for 100 eV primary electrons. Unfortunately, the research did not examine any angular dependence for this effect, and neither, to our knowledge, has subsequent research (Cimino et al., 2004). Nevertheless, we admit the possibility of elastic reflection of incident electrons. We use the following equation to describe the probability with which an electron with energy  $V_i$  is elastically reflected from the channel wall (Cimino et al., 2004):

$$\delta_{el} = \frac{(\sqrt{V_i + V_0} - \sqrt{V_i})^2}{(\sqrt{V_i + V_0} + \sqrt{V_i})^2} \quad (6)$$

where  $V_0$  is an unknown parameter chosen to best fit the data. We find that a value  $\sim 165$ – $175$  eV fits our data well. Eq. (6) is sampled to determine the probability that an electron is elastically reflected after a collision with the channel wall. If reflection occurs, then the axial and angular components of the electrons' velocity are left unchanged, while the radial component is reversed. If reflection does not occur, the secondary electron yield and initial properties of the secondary electrons are determined, as described above.

Channel gain saturation, arising from the presence of large numbers of electrons in the channel, and the build-up of positive wall charge is also included in the model. The effect of the electrons in the pore is approximated by estimating the electrostatic potential due to the electrons cascading in the pore. The MCP channel is divided into some number of axial

segments. In general, we take the segments to be on the order of the average axial distance travelled by the electrons before impacting the channel wall. For a  $D = 10 \mu\text{m}$  pore MCP this distance is  $\sim 10 \mu\text{m}$ . The cloud of electrons in each segment can be approximated as a disc of negative charge, and the electrostatic potential calculated accordingly. The effect of this potential is twofold. First, radial fields exert a force on secondary electrons emerging from the channel wall. When the number of electrons in the channel is high, the exerted force can shorten the electrons' time of flight, thus reducing the time that the axial field accelerates the secondary electrons. The electrons therefore impact with less energy, reducing the average secondary electron yield from the collision. The electron density distribution is constantly changing such that a precise calculation is too complex to be tractable; therefore, the effect for each generation of new secondary electrons is approximated. The second effect that arises is due to axial fields that act as perturbations to the applied accelerating voltage. Generally, these perturbations reduce the accelerating field's strength. This, in turn, reduces the average secondary yield.

As electrons are pulled from the channel walls, a net positive charge remains. The time scale over which the extracted charge is replenished by the current flowing through the MCP is on the order of milliseconds because the lead glass is essentially an insulator. This time scale is much longer than either the  $\sim 180\text{--}220$  ps electron transit time or the subnanosecond gating voltage pulse duration, and so is unimportant for our simulations. The effects of the build-up of positive charge in the channel wall (or wall charging) are similar to those of space charge. Radial fields affect the time of flight of secondary electrons that are emitted into the channel in a way that reduces the gain of subsequent generations. Also, the axial fields perturb the applied fields, thereby altering the trajectories of the electrons so that the gain of secondary electrons diminishes. Furthermore, because the wall charge is neutralized over such long time scales, its build-up can affect subsequent cascades in a given channel, should any occur, and thus has consequences for the linearity of the detector over long exposures, i.e., of order nanoseconds or microseconds. In the work presented here, we are only concerned with a single cascade in a given channel.

Our model simulates MCP response to both static and pulsed voltage biases. Because the value of the voltage does not change over the duration of the electron cascade, static voltage bias is easily understood. We assume, following the work of Gatti et al. (1983), that the electric field is parallel to the channel axis for static voltages rather than perpendicular to the MCP face. According to Gatti et al. (1983), the axial field direction results from the azimuthal current flow around the channel wall's diameter. This current flow ultimately rotates the electric field from a near-perpendicular orientation when the voltage is first applied to an orientation parallel to the channel axis after some milliseconds. In contrast to static bias voltages, for simulations using subnanosecond pulsed voltages, the field is taken to be perpendicular to the face of the MCP.

Time-dependence of the voltage pulse is approximated within the simulation as follows: when a secondary electron is created, the value of the voltage at the time of creation (understood to be the same as the time of the collision of the parent electron with the channel wall) is determined and the electron's trajectory is calculated using that voltage. This approximation should be reasonable if the time scale over which the voltage changes is less than the 5 to 10 ps time of flight of a typical electron in the channel. This assumption holds for the voltage pulses we investigate in this chapter.

The effects of pore end spoiling, where the thin metal layers coated onto the MCP penetrate a small distance into the pore (typically 1–1.5 pore diameters), are also approximated in the

model by setting the field to zero in this region for static and pulsed bias. The zero-field approximation agrees with the work of Price & Fraser (2001), who performed two-dimensional electric field calculations for a single channel's end-spoiled region, showing that the field is close to zero in the end-spoiled region. Landen et al. (2001) show that at high-voltage biases fringing fields in the end-spoiled region may be significant enough that the zero voltage approximation may cease to be valid, and cite experimental results suggesting this is the case. However, the effect primarily occurs when the flux of x rays or UV photons are at a steep angle of incidence relative to the MCP surface. Because the UV and x-ray sources were normal to the MCP surface in the experiments we will later describe, we retain the zero-field approximation in the end-spoiled region.

This completes our description of the Monte Carlo simulation model. The remainder of this chapter discusses simulation results and compares them to experimental data.

### 3. MCP sensitivity under DC voltage bias

Although these simulations can be adapted to MCPs with almost any geometry, we have concentrated most of our modelling efforts on an MCP with a 10  $\mu\text{m}$  channel diameter and a thickness of 0.46 mm ( $L/D = 46$ ). The simulation parameters are discussed in Section 2. A large amount of experimental data exists for these types of MCPs and can be used for comparison. Fig. 2 shows simulation results for an MCP with a steady-state bias of  $-1000$  V. The simulation was initiated by introducing five electrons near the MCP input end. The gain histogram for 2000 separate runs shows that the average gain is about  $3 \times 10^4$ , but a considerable spread in the gain clearly exists. This is a consequence of the statistical nature of the secondary emission process. There is a clear indication of a peak in the gain histogram near 15,000. The transit time distribution for these runs looks essentially Gaussian, with a mean transit time of 188 ps and a full-width half-maximum (FWHM) transit-time spread (TTS) of 53 ps. We lack transit-time measurements for the MCPs we are simulating (such measurements are difficult to make), but these transit times are consistent with existing measurements (Ito et al., 1984).

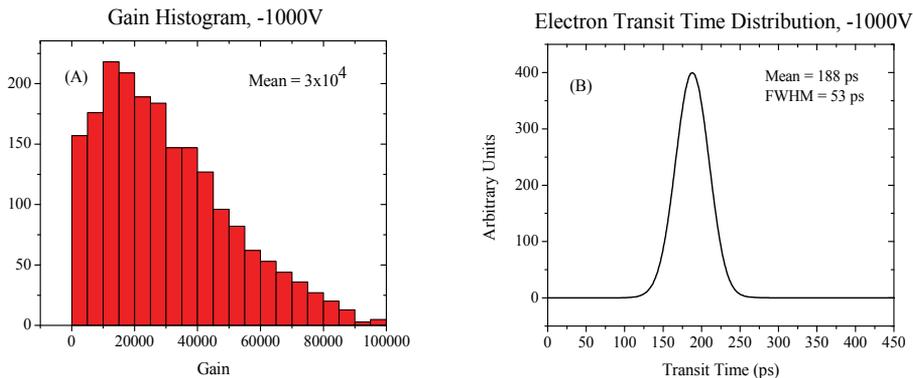


Fig. 2. Simulation results for an  $L/D = 46$ , 10  $\mu\text{m}$  pore diameter MCP biased at  $-1000$  V DC. (A) Gain histogram; (B) electron transit-time distribution.

Fig. 3 shows simulation results for an identical MCP, but with an applied voltage bias of  $-600$  V DC. The average gain is 121, more than two orders of magnitude lower than for the

-1000 V DC simulations, and the gain histogram lacks any indication of a peak. The lower gain, the result of smaller electron impact energies, results in smaller secondary electron yields. The mean transit time is 219 ps,  $\sim 30$  ps longer than for the -1000 V DC simulations. This is a consequence of the decreased acceleration in the -600 V DC bias case. The electrons travel a shorter distance down the channel between collisions, and thus, require more time to reach the output end. Also, the TTS for the -600 V case is 73 ps, much longer than the -1000 V DC case, primarily because the spread in secondary electron energy and direction play a greater role at lower bias voltages where the electrons travel shorter distances between collisions and impact the channel wall with lower energy.

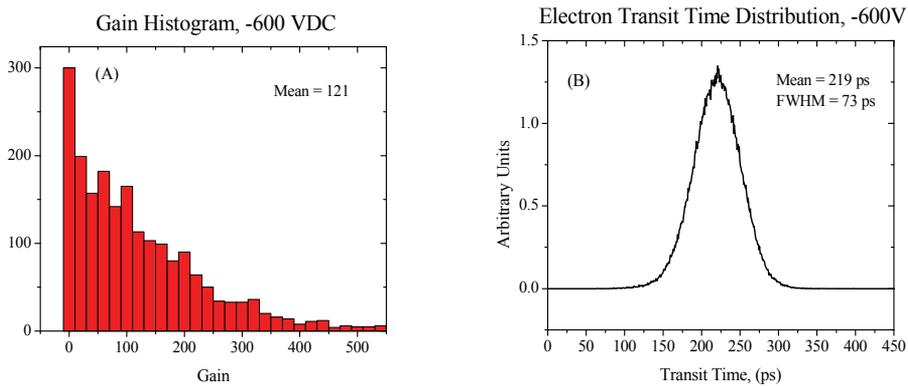


Fig. 3. Simulation results for an  $L/D = 46$ ,  $10 \mu\text{m}$  pore diameter MCP biased at -600 V DC. (A) Gain histogram; (B) electron transit-time distribution.

In order to check the validity of our simulation results, we compared the modelled versus measured MCP sensitivities. The experimental details of the MCP DC sensitivity measurements using a Manson x-ray source were described previously (Wu et al., 2008). The detector consists of an MCP coated with six separate strips (each 4 mm wide  $\times$  40 mm long, separated by 2 mm) on the input (bias) and output (ground) surfaces, a P43 phosphor screen coated on a fiber-optic faceplate at positive potential and a coherent fiber bundle coupled between the faceplate and a charge-coupled device (CCD). A negative-bias voltage was applied to each strip, and the phosphor-coated fiber-optic faceplate was held at +3000 V with respect to the MCP back surface. The Manson source was operated at 8 kV with 0.3 mA of emission current using aluminium anodes that have emission peaks at about 1.5 keV. The x-ray flux and spectrum was monitored by an Amptek XR-100-CZT x-ray pulse-height spectrometer. The MCP were placed  $\sim 2.7$  m from the Manson source to obtain a uniform x-ray flux on all strips. Two Uniblitz x-ray shutters installed on each line of sight ensured equal x-ray exposures. Beryllium filters were used on both lines of sight to block light emission from the filament. Relative sensitivities were measured as a function of voltage for potentials ranging from -450 to -950 V, incremented by 50 V. Three images were taken at each voltage setting. The measured intensity was obtained by averaging all six strips and the error bar was given by one standard deviation of sensitivity over the entire MCP.

Both modelled and measured sensitivities are plotted versus voltage in Fig. 4. The simulated gains have been scaled so that the model value at -450 V is set to the average of the measured data. Clearly the model reproduces the trend in the measured data extremely well over virtually the entire voltage range. However, the data show some levelling off at -950 V,

which is not seen in the simulations. This levelling off is due to the onset of saturation in the CCD used to collect the data, and is not an effect of the MCP. This excellent agreement between the experiment and simulation indicates that the parameters used in the model as described in Section 2 are reasonable. With the success in DC simulation, it gives us confidence to attempt pulsed mode simulations.

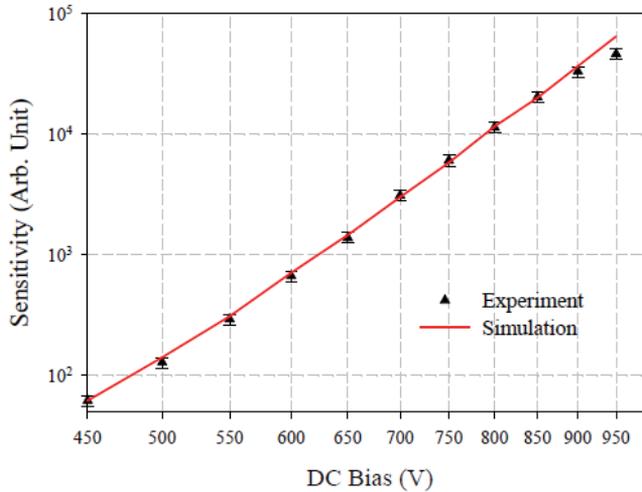


Fig. 4. Comparison of simulated and measured MCP relative sensitivity vs. DC voltage shows that parameters used in our model agree with existing experimental data.

#### 4. MCP performance under an ideal square waveform

Initially, our investigation of pulsed MCP response using our Monte Carlo simulation model was conducted by looking at an ideal square-wave voltage pulse. The ideal square-wave, though only an approximation of a real voltage waveform, offers useful insights about expected MCP behavior under voltage pulses of varying widths and peaks. The simulations used a  $-700$  V pulse and a DC offset varied from  $+400$  V to  $-500$  V in increments of  $50$  V so that the peak voltage varied from  $-300$  to  $-1200$  V. The pulse widths were varied from  $50$  to  $300$  ps in  $50$  ps increments. Five primary electrons initiated the cascades.

The peak MCP gain as a function of peak voltage plotted on a log-log scale is shown in Fig. 5. A few observations can be made about these results. First, it is apparent that for pulse widths shorter than the  $\sim 200$  ps transit time, there is essentially no gain for a reverse DC bias. This is unsurprising, since very few electrons make it through the MCP in such a short time, and the direction of the DC field acts to prevent any electrons emerging from the output end once the pulse has stopped. Second, it is evident that the relative gain as a function of peak voltage varies greatly for different voltage pulse widths. For pulse widths  $< 250$  ps there is a larger gain exponent ( $n$  in the expression  $G \sim V^n$ ) as a function of peak voltage than there is for  $250$  and  $300$  ps pulses, which give a gain exponent of  $n = 10.8$ , close to the  $n = 11$  DC mode gain exponent. Since these are ideal square waveforms, it is expected that the sensitivity of gain to peak applied pulsed voltage should be nearly identical to the DC results when the pulse length is longer than the electron transit time, as the MCP is

essentially operating in DC mode for such pulses. The primary difference between the longer pulses and DC mode operation is the slightly different electric field configuration referred to in Section 2. However, the gain behavior of the MCP departs substantially from the DC mode when the pulse width is smaller than the transit time of electrons. The shorter the pulse width the stronger the deviation from the DC mode will be, as shown in Fig. 5. It is evident that when the applied voltage pulse width becomes shorter than the transit-time, it is no longer possible to obtain the full gain from the MCP during the pulse. In this case, the gain of detector has a higher order of nonlinearity than in DC mode.

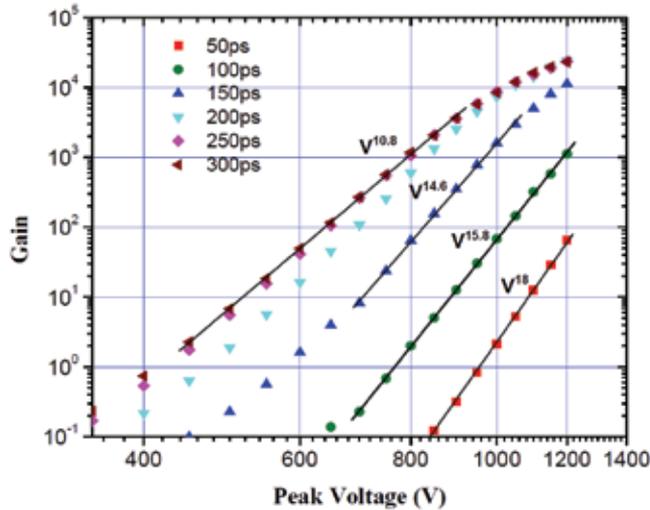


Fig. 5. Simulated peak gain vs. peak voltage for an ideal square wave. The result indicates that shorter pulse widths deviate most strongly from DC mode.

The increase in the gain exponent for pulses shorter than the transit time has previously been reported by Landen et al. (1993), who described data giving a gain exponent of  $n = 20$  for voltage pulses shorter than the transit time of the electron cascade through a similar ( $L/D = 46$ ) MCP. From our simulations, a gain exponent of  $n = 18$  is estimated for the 50 ps ideal square pulse, somewhat smaller than they report. It is difficult to make a direct comparison between the Landen et al. (1993) results and these simulations because they made no measurements of the actual voltage waveforms on the MCP. Thus, the precise time dependence of the voltage pulses is unknown.

These simulation results suggest a limitation to how fast one can gate the MCP. For a particular MCP, when the width of the applied voltage waveform is less than the average transit time, the detector's gain will be reduced significantly. A gain reduction of more than three order of magnitude is calculated when the pulse width is reduced from 250 to 50 ps with a peak voltage at 1000 V. This is not a surprising result: a shorter voltage pulse means fewer electrons transit the MCP and fewer collisions with enough energy to produce secondary electrons occur. However, it offers further illustration of the difficulty of gating an MCP with pulses comparable to and shorter than the transit time. The maximum voltage that could be applied to the  $L/D = 46$ , 10  $\mu\text{m}$  pore MCP is most likely limited to be about -1200 V to avoid breakdown. The MCP's gain would be  $<100$  for a 50 ps square pulse with a -1200 V peak voltage. As described above, for pulse widths shorter than the transit time the

MCP exhibits a larger gain exponent. Therefore, for such voltage pulses the relative detection sensitivity of the MCP follows a higher power law. In other words the MCP is much more sensitive to peak voltage. Thus, slight variations in the peak voltage along the MCP, which can result from attenuation of the voltage pulse along the MCP strip and reflections from the strip ends, will have a much greater impact on the detection uniformity along the MCP strip for pulse widths shorter than the transit time.

## 5. MCP performance optimization in pulsed mode

The Monte Carlo simulation code can also be used to predict detector performance, providing guidelines for selecting detector operating parameters. An  $L/D = 46$  MCP is used here as an example of how to employ our Monte Carlo approach to achieve a narrow optical gate profile without significantly reducing detection sensitivity. The CPS3 pulser system from Kentech Instruments Ltd. is widely used in the high-energy density physics (HEDP) community to drive gated MCP detectors. Using a measured waveform from a CPS3 pulser and a 300 ps near-square-top pulse-forming network (PFN), we cut 50 ps increments from the ‘flat’ section of the waveform, thereby constructing hypothetical waveforms for use in the simulation code with flat tops of 300, 250, 200, 150, 100, 50, and 0 ps. The flat-top pulse width for the input waveforms here is defined as the width of 95% of peak voltage, while the 0 ps flat-top waveform is simply a combination of the rising and falling edges of the measured pulse. The peak voltage of the waveforms was scaled to be near 800 V; DC transit times at 800 V are near 200 ps. Seven of the input waveforms and simulated MCP gate profiles are shown in Figs. 6(A) and (B). Significant changes in the FWHM of the gate profiles and the MCP relative gain are observed. The FWHM of the gate profiles vs. the pulse width of the flat top is shown in Fig. 7(A). Gate profile FWHMs are 230, 186, 150, 115, 105, 99, and 97 ps, respectively, for the 300, 250, 200, 150, 100, 50, and 0 ps flat-top waveforms. Of note is that the limiting temporal resolution, as defined by the gate profile FWHM, for the  $L/D = 46$ , 10-micron pore MCP appears to be about 100 ps regardless of the voltage pulse width.

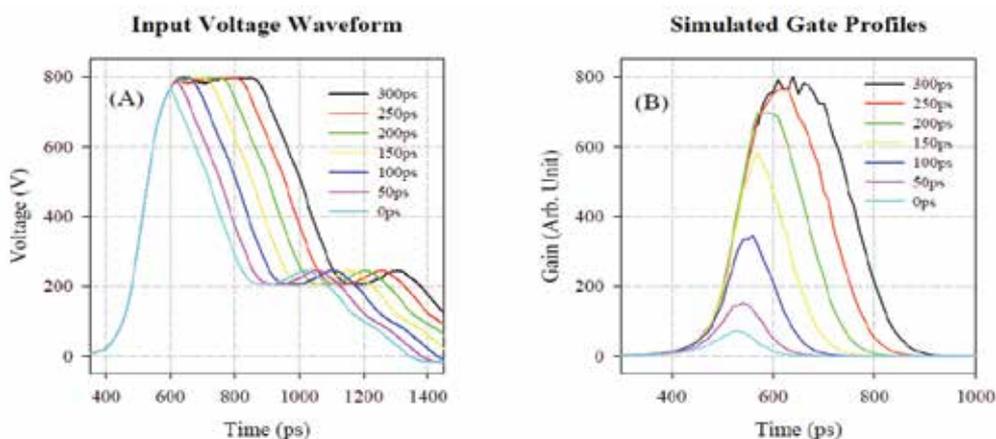


Fig. 6. (A) Input voltage waveforms and (B) Simulated optical gate profiles for 10  $\mu\text{m}$  and  $L/D = 46$  MCP.

The integrated gate profiles and relative peaks of the gate profiles with these voltage waveforms are shown in Fig. 7(B). The results show that the overall detection sensitivity, defined as the integrated gate profile, will be reduced by about 60% when the flat top of waveform is reduced from 300 to 150 ps, while the gate profile peak amplitude changes only about 25%. The overall detection sensitivity will be reduced by about a factor of ten when a 50 ps flat-top waveform is applied as compared with the 300 ps flat-top waveform. Depending on specific experimental requirements, a compromise between the detection sensitivity and temporal resolution is necessary. These results indicate that when the 150 ps flat-top waveform is selected, the predicted gate profile width, 120 ps FWHM, is close to the shortest achievable gate profile. At the same time, with only a 25% reduction in the peak gain, relatively little is sacrificed in the way of detector sensitivity. Thus the 150 ps flat-top PFN would seem to offer a reasonable compromise. In the following section we discuss measurements made using a 150 ps flat-top PFN and comparisons to further simulations.

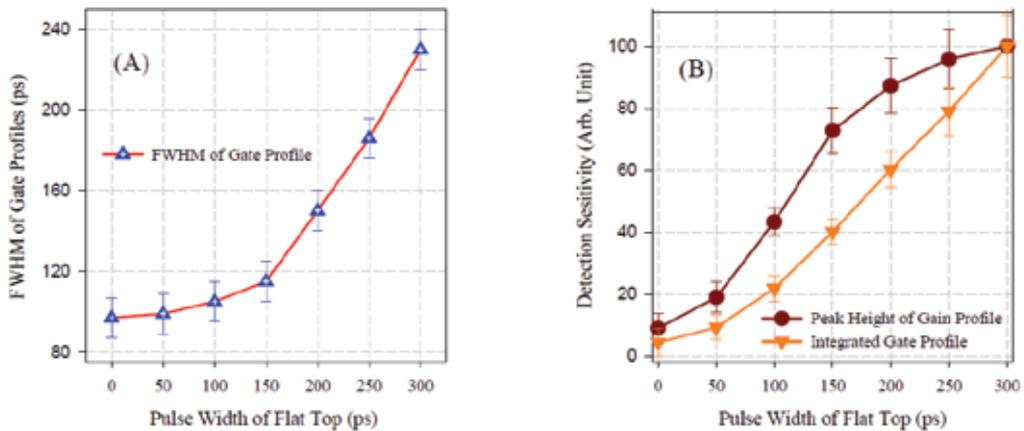


Fig. 7. (A) Variation of FWHM of the simulated gate profiles versus waveform flat-top width; (B) detection sensitivity versus waveform flat-top width.

## 6. Optical gate profiles and detector uniformity

Optical gate profiles and detector uniformity along the MCP strip are two important parameters for HEDP diagnostics applications. The ability to use simulations to predict the performance of a time-gated MCP detector is highly desired in the HEDP community, because of the time and high cost involved in characterizing detector performance experimentally. Because the detector gain with applied voltage in pulsed mode is highly nonlinear, slight variations in the voltage waveform on the MCP strips can significantly change their performance. The actual voltage waveform on each MCP strip on each detector can be affected by assembly processes and component quality, even for detectors of the same design. For high quality measurements, it is necessary to characterize the performance of each individual detector. Once we are able to verify our Monte Carlo code and the MCP physical model by comparison to experiment, the simulation code can be used to predict the performance of the MCP detectors when actual waveforms on the MCP strips are precisely measured.

Experimental details of the characterization process have been published before (Kruschwitz et al., 2008; Rochau et al., 2008) and will only be described briefly here. MCP detector characterizations were conducted at the Short-Pulse Laser Facility at National Security Technologies, LLC, which provided 200 nm laser light with a 150–200 fs pulse width at 150–200  $\mu\text{J}$  energy per pulse. The laser beam was expanded to cover the entire MCP detector, and uniformity of the laser beam was achieved using a homogenizer and diffuser. The laser flux was adjusted with neutral density filters to ensure that the MCP was not saturated. A coherent fiber plug was used to couple the phosphor and the CCD camera.

The detector used in the experiment is the NSTec H-CA-65 camera shown in Fig. 1, an eight-frame, gated MCP framing camera that uses a rectangular MCP  $42 \times 48$  mm in size with a  $38 \times 42$  mm active area. The MCP is rimless, has 10  $\mu\text{m}$  diameter pores, a pore length of 460  $\mu\text{m}$ , and an open area ratio of 65%. Eight independent 4 mm wide striplines, comprised of 5000  $\text{\AA}$  of Cu with an overcoat of 1000  $\text{\AA}$  Au, are coated onto the MCP with an impedances of  $\sim 19 \Omega$ . The internal circuitry consists of input and output flexible circuit boards. The input circuit board has 20  $\Omega$  transmission lines, while the output circuit board has 20 to 50  $\Omega$  transmission line tapers to prevent any reflections of the input high-voltage (HV) pulse. The MCP is gated using a pulse from a Kentech CPS3 pulser and a PFN that provides an approximately 150 ps flat-top (450 ps FWHM) voltage pulse with a peak amplitude of approximately  $-1500$  V at the input of the detector. Impedance mismatches at the detector input reduce the peak voltage on the MCP to approximately half of that value because only about  $-800$  V pulse waveform is needed to meet our experimental conditions.

Experimental optical gate profiles were obtained using the following procedure. MCP detector images were recorded for a series of time delays between the laser and the HV pulse on the MCP. The timing of the laser was measured by a fast photo-conductive diode (PCD), while the timing of HV pulse was given by the CPS3 output monitor. Five images were taken at each time delay. The time delays are determined from the relative timing between the PCD signal and the voltage pulse as measured on a 16 GHz, 50 GS/s oscilloscope. The 50% points of the rising edges on each of these signals could be determined to  $<10$  ps. The timing jitter in the experiments was about  $\leq 50$  ps. The gain at a specific time and location on the MCP striplines is determined using the following procedure: (1) Subtract the CCD background from the image; (2) Divide the background subtracted image by the flat-field of the beam profile taken with an applied DC voltage; (3) Integrate a narrow region around the spatial location of interest to get the average intensity; (4) Divide the average intensity by the measured laser energy. (5) The gate profiles are obtained by sorting the averaged intensities of locations of interest by the measured time delays with a bin width of 10 ps. The position-dependent gate profiles along the MCP strip are given in Fig. 8 (B).

Due to the non-linearity of gain with applied voltage, it is critical to precisely measure the voltage waveforms on the MCP strip in order to make a meaningful comparison between the simulations and experiments. The details of the experimental setup and procedure for making these measurements have been described previously (Rochau et al., 2008) and will only be summarized here. A high-impedance GGB Industries Model-35 Picoprobe is used to measure the time dependent voltage waveform at various spots along an MCP stripline. The probe has a 20  $\mu\text{m}$  tungsten wire tip, an input impedance of 1.25 M $\Omega$ , and a frequency response from DC to 26 GHz. The probe is mounted on an x-y-z station controlled by step motors, so it can move to any desired location on the strip with high accuracy. The voltage

waveforms at different positions along a particular MCP strip, as measured with the Picoprobe, are shown in Fig. 8(A). The input voltage waveform for the detector is a near-square waveform, but the waveform on the MCP strip is quite different, with essentially no flat top visible. Also, transmission loss on the strip can be significant and can cause the voltage waveform to vary along the strip. The amplitude of the voltage pulse at the strip's center is less than at beginning of the strip due to attenuation of the voltage pulse along the MCP strip. However, the voltage pulse amplitude at the end of strip is larger than at the center because the voltage pulse is reflected at the junction between MCP strip and the circuit board where there is a slight impedance mismatch.

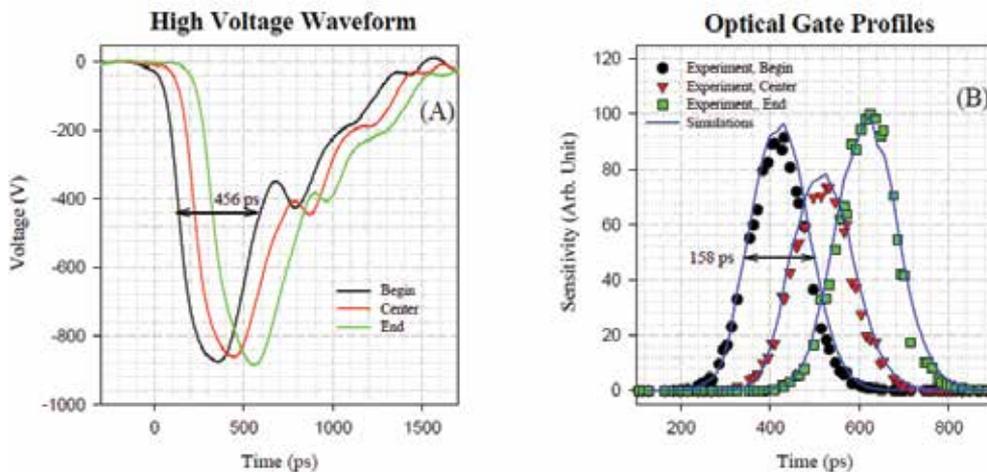


Fig. 8. (A) Measured waveforms at different positions on the strip; (B) measured and simulated gate profiles at various positions along the strip.

The waveforms in Fig. 8(A), used as input for the Monte Carlo model, allowed us to simulate the position-dependent gate profiles by beginning electron cascades at different times during the voltage pulses. In effect, this simulates the different time delays between the short-pulse laser and the HV pulse so that a valid comparison can be made between the simulated and measured profiles. Fig. 8(B) shows such a comparison. The agreement between the measured and simulated gate profiles is excellent. Also notable is that the FWHM of the gate profiles along the entire MCP are within the experimental error of  $\pm 10$  ps, so the gate profile width is essentially position independent, which is important for HEDP diagnostics applications. There is, however, an obvious change in the gate profile peak, with the peak at the center of the strip about 20–25% smaller than at either end of the strip; this results from voltage pulse attenuation along the strip.

Detector flat-field characterization in pulsed mode provides an interesting comparison between the simulations and experimental measurements. Generally, MCP detector flat-field characterization in pulsed mode is done using a long-pulse x-ray source or UV laser, with the source pulse lasting much longer than the voltage pulse duration. It is also possible to obtain the flat-field of the MCP detector from a short-pulse UV laser by integrating the position-dependent gate profiles. In Fig. 9 sensitivity uniformity along the MCP strip between experimental results and simulations are compared. The integrated intensity of the experimental data here is simply a sum of the position-dependent gate profiles, normalized

for the entire MCP strip. The integrated intensity obtained from the simulations is scaled to the experimental data at 33 mm. The differences between experimental and simulation data are within their error bars. The experimental error bar is a standard deviation within each spatial location of interest, while the error bars on the simulations are taken to be 10%.

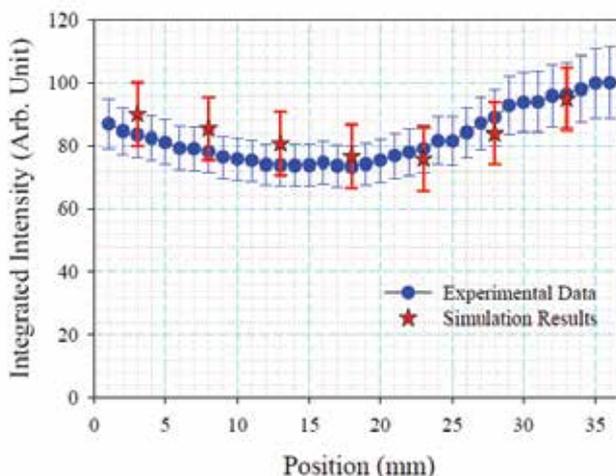


Fig. 9. Simulated and measured detection uniformity along the MCP strip.

## 7. MCP gain in pulsed mode

To make a comparison to the square wave results shown previously and to the results of Landen et al. (1993), experiments were also conducted to examine how the MCP gain sensitivity varied with peak applied voltage. The voltage pulse from our CPS3 unit with a 150 ps flat-top PFN plus various DC offsets were used to bias the MCP. The delay timing was set so that the laser and HV pulses overlapped at strip's center. Due to the timing jitter, however, the maximum signal was often not at the center of the strip. Although five images were taken, timing jitter was the primary source of error.

For comparison with the experimental results, we ran simulations using the voltage profile shown in Fig. 8(A) for the center of the strip. The electric field due to this voltage pulse was added to the electric field from the measured DC offsets, and electron cascades were launched at the appropriate times to achieve the maximum gain for that particular voltage pulse. An average of five electrons was used to initiate the cascades. To account for uncertainties in the true voltage value on the plate, we incorporated a  $\pm 5\%$  uncertainty in the peak voltage pulse value. The results are shown in Fig. 10 along with the experimental data. The error bars on the simulation data points are calculated based on the assumed  $\pm 5\%$  uncertainty in the peak of the voltage pulse and a one-sigma uncertainty calculated from the simulated gain distribution. From Fig. 6(A), it seems that the variation of peak voltages along the MCP strip is less than  $\pm 5\%$  and so the errors estimated for the simulation are likely an upper limit. Power dependences of 16.9 and 15.4 were observed for the experiments and simulations, respectively. The agreement between results is within the error bars. The higher order power dependence of the relative gain on peak voltage when the flat top of the HV waveform is less than the transit time of cascading electrons in MCP is thus evident.

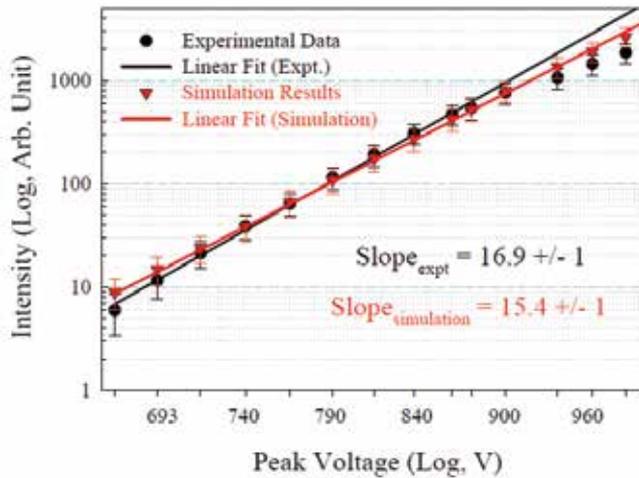


Fig. 10. Simulated and measured relative sensitivity vs. peak voltage.

## 8. Gain saturation in DC and pulsed modes

It is essential to understand the saturation limit when considering experimental application of MCP detectors. In this section, to examine saturation behavior of MCPs under both DC and pulsed modes, the average intensity of the MCP was calculated over entire MCP strips at each laser flux, while the experimental errors were estimated by the standard deviation of the sensitivity. In the pulsed mode, the delay timing allowed the laser and HV pulses to overlap at the center of the MCP strip. Due to the timing jitter, the maximum signal was often not at the center of the strip; the average peak intensity was a mean value of five maximum intensities over the strips in each selected laser flux, which were also normalized with fluctuation of the laser power. The experimental errors were standard deviations of intensity with the bin size (1 mm) and five laser shots.

In the simulations, the effect of increasing the laser power was approximated by increasing the mean number of primary electrons. A true comparison to the experimental data would require knowledge of the quantum efficiency of the MCP for the 200 nm photons produced by the laser, which is not available in literature. However, given the large number of photons/channel per laser pulse for even the lower fluxes we conclude that it must be about  $\leq 0.1\%$  for the MCP to see no indication of saturation at these fluxes. The comparison of simulations to data is based on the relative increase in primary electron number, assuming that the mean number of primary electrons scales linearly with the laser flux. Thus, we performed simulations with the mean number of primary electrons  $\delta_0$  varying between 100 and 5000 for DC bias voltages between  $-400$  and  $-900$  V at 50 V steps. Also, to investigate the detector dynamic range in pulsed mode, we performed simulations using a pulsed waveform measured by the Picoprobe, and with the number of primary electrons,  $\delta_0$ , varying between 1 and 3000.

DC mode experimental data and simulation results have been plotted together in Fig. 11. The experimental data show appreciable saturation at  $-700$  VDC for photon fluxes  $>500$  nJ/cm<sup>2</sup>, but saturation at lower fluxes is absent. It is also true that the gain versus voltage sensitivity changes somewhat with laser power. As the laser flux is increased, the gain

becomes slightly less dependent on voltage, changing from about a  $G \propto V^{11}$  dependence to more nearly a  $G \propto V^9$  dependence before the onset of saturation. The simulations exhibit a similar trend, changing from  $G \propto V^{11}$  at a mean of one primary electron to  $G \propto V^{8.5}$  in the linear (unsaturated) range at a mean of 5000 primary electrons. This effect, which was first reported in Kruschwitz et al. (2008), implies that this decrease in the gain sensitivity may arise from a “weak” saturation resulting from high electron numbers present in the channel at a given time, even though the MCP gain may not be very high, because the photoelectrons are generated in less than 1 ps by the laser pulse. In other words, the “weak” saturation may be a space charge effect due to a high production rate of photoelectrons.

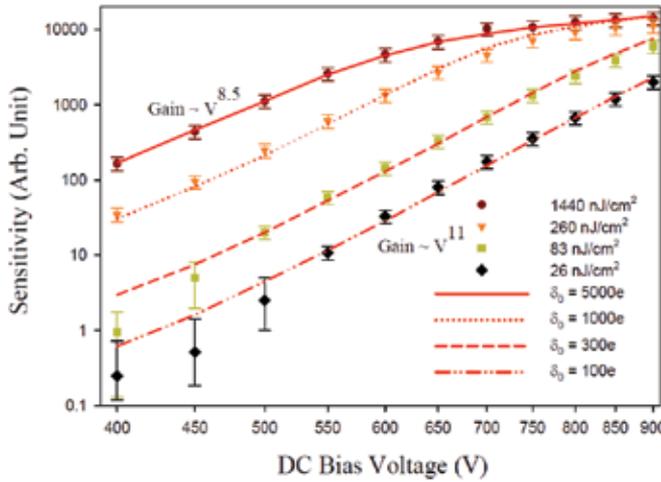


Fig. 11. Simulated and measured MCP sensitivity vs. DC voltage. Data were obtained using a UV laser and neutral density filters to adjust the flux. Simulations were run with different numbers of initial electrons.

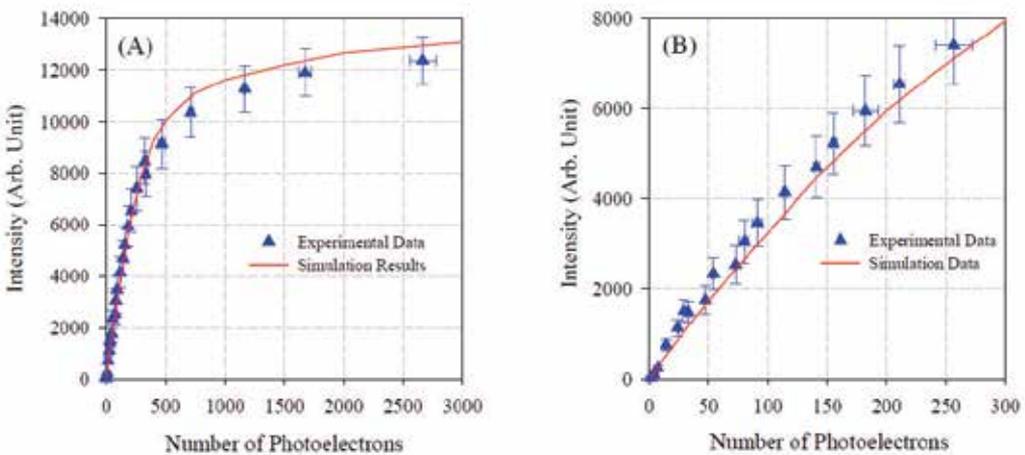


Fig. 12. Simulated and measured MCP saturation behavior in pulsed mode.

Fig. 12 shows a comparison of simulated and measured MCP saturation in pulsed mode. The measured laser flux was converted to a number photoelectrons per pore assuming the MCP's quantum efficiency at 200 nm wavelength is 0.1%, since the absolute detection efficiency of the MCP detector and recording system were not well determined. The number of output electrons was then scaled to the measured output signal. Fig 12(A) shows good agreement for entire range from 1 to 3000 input photoelectrons. A low photon flux range is expanded in Fig. 12(B). The differences between the experimental data and simulation results are well within the experimental error bars. But the simulation results seem in general to be a bit lower in the lower flux range and somewhat higher in the saturation region. It is also interesting that, according to simulations, the saturation limit is about 300,000 electrons per pore, and the onset of nonlinearity is near 100,000 electrons per pore for a 10  $\mu\text{m}$  pore diameter,  $L/D = 46$  MCP. When the MCP gain is set at 100 and assuming a detection limit of 10 electrons, the dynamic range of the MCP could be on the order of just a few hundred.

### 9. Spatial resolution

A useful characteristic of the three-dimensional electron transport model is that it allows us to predict MCP resolution. Resolution calculations were based on the set of parameters for the MCP camera back imaging system described in the experiments. As described above, the electron position distribution is obtained by calculating the ballistic trajectory of the MCP output electrons hitting the phosphor plate held at a fixed positive potential. The electron scattering from the phosphor surface does not include in the simulation model. The phosphor is located 0.75 mm. from the MCP exit face, which is at ground potential. Experimental measurements of the detector spatial resolution were made using a knife-edge resolution target. For some simulations, we assumed that the conductive gold coating extended into the channel output a distance of about 1.5 channel diameters, or 15  $\mu\text{m}$ . This end spoiling created an electron focusing effect, which has been detailed previously (Eberhardt, 1979; Bronshteyn et al., 1979; Koshida & Hosobuchi, 1985; Koshida, 1986).

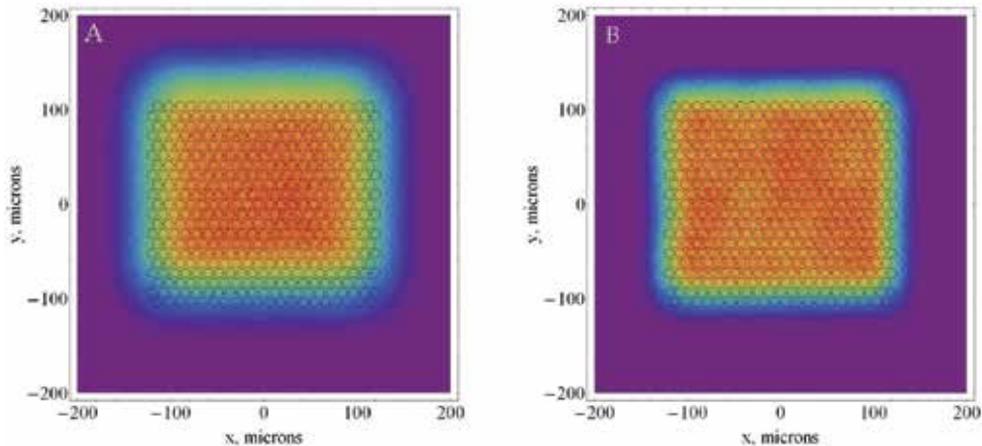


Fig. 13. Simulated output electron distributions on phosphor: (A)  $V_{ph} = 1000$  V; (B)  $V_{ph} = 4000$  V for 10  $\mu\text{m}$   $L/D = 46$  MCP with a 0.75 mm gap between the MCP and phosphor.

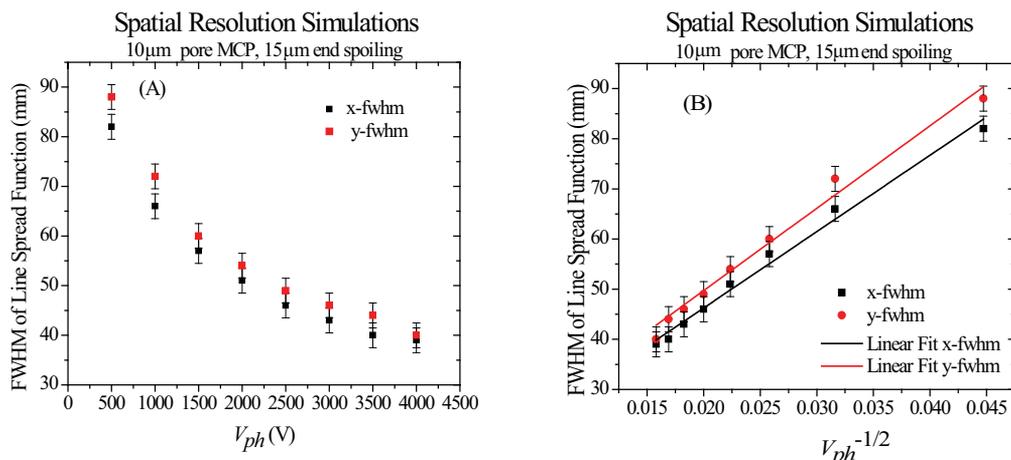


Fig. 14. Spatial resolution vs. applied voltages (A)  $V_{ph}$  and (B)  $V_{ph}^{1/2}$  on the phosphor with 15  $\mu$ m of end spoiling.

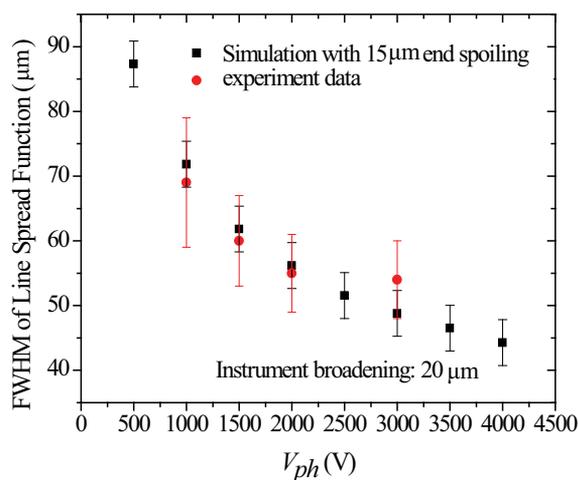


Fig. 15. A comparison between measured and simulated spatial resolution.

For our spatial resolution calculations, we have generated a set of output electrons from a 430-channel array in a hexagonal-packed geometry with a center-to-center distance of 12  $\mu$ m. Electron cascades were launched in each pore to approximate a two-minute exposure on the Manson x-ray source. An orthogonal x–y coordinate system was established in the phosphor plane with the bias angle along the y axis. Fig. 13 shows a simulated image of the output from the 430-pore array; locations of the pore output ends are shown. 15  $\mu$ m of pore end spoiling was assumed. The two images are for  $V_{ph} = 1000$  V and  $V_{ph} = 4000$  V. Line spread functions (LSF) were calculated by taking the first derivative of the electron distribution functions, such as those shown in Fig. 13. The FWHMs of the LSFs were calculated for phosphor potentials  $V_{ph}$  ranging from 500 to 4000 V. The resolution in the x and y directions, defined as the FWHM of the LSFs, are plotted in Fig. 14. In Fig. 14(A) they are plotted simply as function of  $V_{ph}$ , while in Fig. 14(B) they are plotted as a function of  $V_{ph}^{-1/2}$ .

<sup>1/2</sup>. A simple analytical expression for the spatial resolution  $\delta$  of an MCP detector indicates that the relationship between  $\delta$  and  $V_{ph}$  is  $\delta \propto V_{ph}^{-1/2}$  (Kilkenny, 1991) and, so, a plot such as that in Fig. 14(B), should yield a straight line. From Fig. 14(B), we see that such a relationship does in fact hold, in the x and y directions. It is also evident from Fig. 14 that the LSF FWHM in the y direction is slightly larger than in the x direction.

Fig. 15 compares the experimental measurements of the spatial resolution of the MCP plus CCD detector system and the simulations of the MCP spatial resolution. Experimental measurements and simulations have been performed for  $V_{ph}$  between 500 and 4000 V. For the simulated resolutions, 20  $\mu\text{m}$  instrument resolution broadening was added in quadrature to the simulated LSF FWHM with 15 microns of end spoiling, and the x and y FWHM were averaged. The measured and simulated results show an excellent agreement.

## 10. Simulations of small-pore MCPs

An important aspect of the Monte Carlo model is its adaptability. It is a relatively simple matter to investigate MCPs with a broad range of parameters. Small-pore MCPs, with pore diameters down to 2  $\mu\text{m}$  and a center-to-center spacing of 3  $\mu\text{m}$ , are now available. The Monte Carlo code was used to study such MCPs and their potential characteristics for imaging applications. Potential benefits to using small-pore MCPs would be faster time response, which translates to superior time resolution, and perhaps improved spatial resolution due to the smaller pore size. Drawbacks would be inferior dynamic range and the small sizes of the MCPs (typically  $\sim 1$  cm diameter and thicknesses on the order of 100  $\mu\text{m}$ ), which make them far more delicate and difficult to work with than their larger-pored brethren.

In order to understand potential advantages and limitations of the small-pore MCP, we modelled MCPs with an L/D ratio of 60 and an 8-degree bias angle for 10 and 2  $\mu\text{m}$  pore sizes. Cascades were started with a mean of three electrons. Fig. 16(A) and (B) shows the simulated DC sensitivity versus voltage for L/D = 60, D = 10  $\mu\text{m}$  pore MCP, and D = 2  $\mu\text{m}$  pore MCPs respectively. The gain curves look quite similar, as one would expect given that the L/D ratio is identical for both MCPs. The 2  $\mu\text{m}$  pore MCP begins to show some nonlinearity at slightly less than  $10^4$  electrons versus slightly less than  $10^5$  for the 10  $\mu\text{m}$  pore MCP. Due to the decrease in effective inner surface area between the 2 and 10  $\mu\text{m}$  pores, the saturation limit of the 2  $\mu\text{m}$  pore MCP can be reduced by more than a factor of ten if surface charge depletion is the dominant factor. The small-pore MCP's dynamic range can thus be expected to be reduced by about a factor of ten or more compared to the 10  $\mu\text{m}$  pore MCP. Fig. 16(C) and (D) show the transit time distributions for the same MCPs. As shown, the 2  $\mu\text{m}$  pore MCP has a much shorter transit time (55–60 ps) and a much narrower TTS than the 10  $\mu\text{m}$  pore MCP. As shown in Section 9, this implies that one could design an imager with temporal resolution near 50 ps without sacrificing much gain, something much desired in the HEDP community. An attempt to achieve gate times in the 30 – 40 ps range was reported by Bradley et al (1995) using an L/D = 20, D = 10  $\mu\text{m}$  pore MCP available at the time. A -1500 V pulse was applied to that MCP in their calculated estimate of the gate time. That voltage might not be physically possible due to potential arcing through the 220  $\mu\text{m}$  MCP. Our simulation here using L/D = 80 2  $\mu\text{m}$  pore MCP implies that an x-ray imager with an optical gate near 50 ps can potentially be realized with currently available technology (i.e., MCPs and HV pulsers).

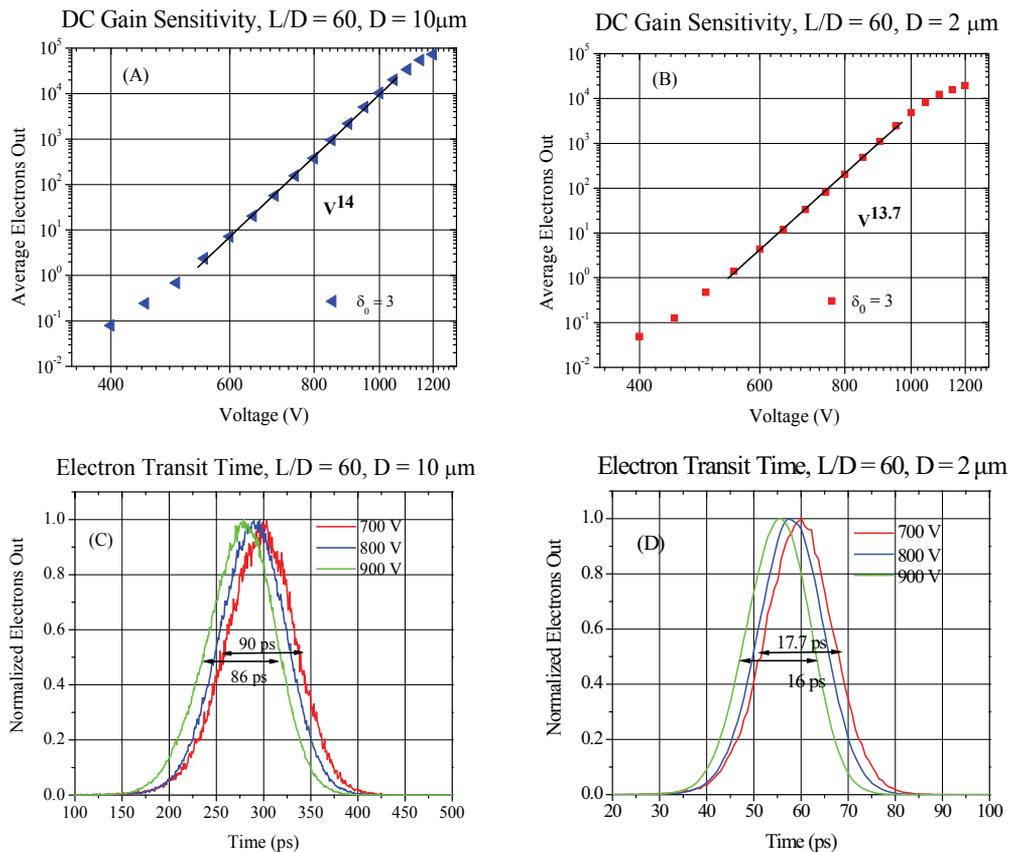


Fig. 16. A comparison of DC sensitivity (A&B) and transit time and TTS (C&D) between 10 and 2  $\mu\text{m}$  L/D ratio of 60 MCPs by the simulations.

Simulations of the gate profiles for the 2  $\mu\text{m}$  pore MCPs were also performed using the same setup described earlier for the L/D = 46, D = 10  $\mu\text{m}$  MCPs: we used a measured voltage pulse from a Kentech CPS3 pulser with a 300 ps flat-top PFN, and cut 50 ps portions from the flat region, as shown in Figure 5(A). MCPs with L/D = 46, 60, and 80 were investigated. Simulated gate profiles of these MCPs are shown in Figs. 17(A), (B), and (C) separately. Gate profiles for the shortest voltage pulses are about 55 ps, suggesting the possibility of achieving imaging time resolutions of near 50 ps using readily available technology. Such resolutions are of great interest in the HEDP community. Fig. 17(D) shows relative sensitivity variations in both peak height and integrated area. Because 2  $\mu\text{m}$  MCPs have a shorter transit time, there is not much change in peak height when the flat top is longer than 10 ps. When the optical gate is near 55 ps, the peak height is reduced less than a factor of five. Therefore, if a pulse with a faster rising and falling edge could be achieved, then even shorter time resolutions may be possible.

Our study of small-pore MCPs is completed by investigating the spatial resolution of a hypothetical imaging system using a 2  $\mu\text{m}$  pore MCP. Like the 10  $\mu\text{m}$  pore MCP spatial resolutions simulations, a grouping of 430 pores was studied. A  $d = 0.5$  mm gap between the output face of the MCP and the phosphor was used in most of the simulations, but a few were done with a gap of  $d = 0.25$  mm. One hundred cascades were started in each pore in

the cluster and the output electrons were accelerated by the MCP-phosphor bias voltage,  $V_{ph}$ . Values of  $V_{ph}$  ranging from 500 to 4000 V were investigated for those simulations using  $d = 0.5$  mm, while  $V_{ph}$  values of 500 to 2000 V were used for those simulations using  $d = 0.25$  mm. Fig. 18 shows the simulated electron output for the 430-pore cluster for  $V_{ph} = 1000$  V and  $V_{ph} = 4000$  V,  $d = 0.5$  mm with  $3 \mu\text{m}$  (1.5 channel diameters) end spoiling. The output electron distribution is clearly far more diffuse for the  $V_{ph} = 1000$  V simulations. Fig. 19 shows the FWHM of the LSFs for all of the investigated values of  $V_{ph}$ , plotted both as a function of  $V_{ph}$  and as a function of  $V_{ph}^{-1/2}$ .

There appears to be a linear relationship in both the x- and y-directions between the FWHM of the LSF and  $V_{ph}^{-1/2}$ , at least for voltages greater than 2000 V. Similar to the  $10 \mu\text{m}$  pore MCP simulations, the FWHM in the x-direction is smaller than in the y-direction for  $V_{ph} \geq 2000$  V. In contrast to those results, however, for  $V_{ph} < 2000$  V the FWHM in the y-direction is much smaller than in the x-direction. In fact, the FWHM in the y-direction appears to be independent of  $V_{ph}$  for  $V_{ph} < 2000$  V. For both sets of small-pore simulations the spatial resolution can be expected to be improved relative to the  $10 \mu\text{m}$  pore MCP. The FWHM of the simulated LSFs are smaller by  $\sim 10 \mu\text{m}$  for the  $2 \mu\text{m}$  pore MCPs.

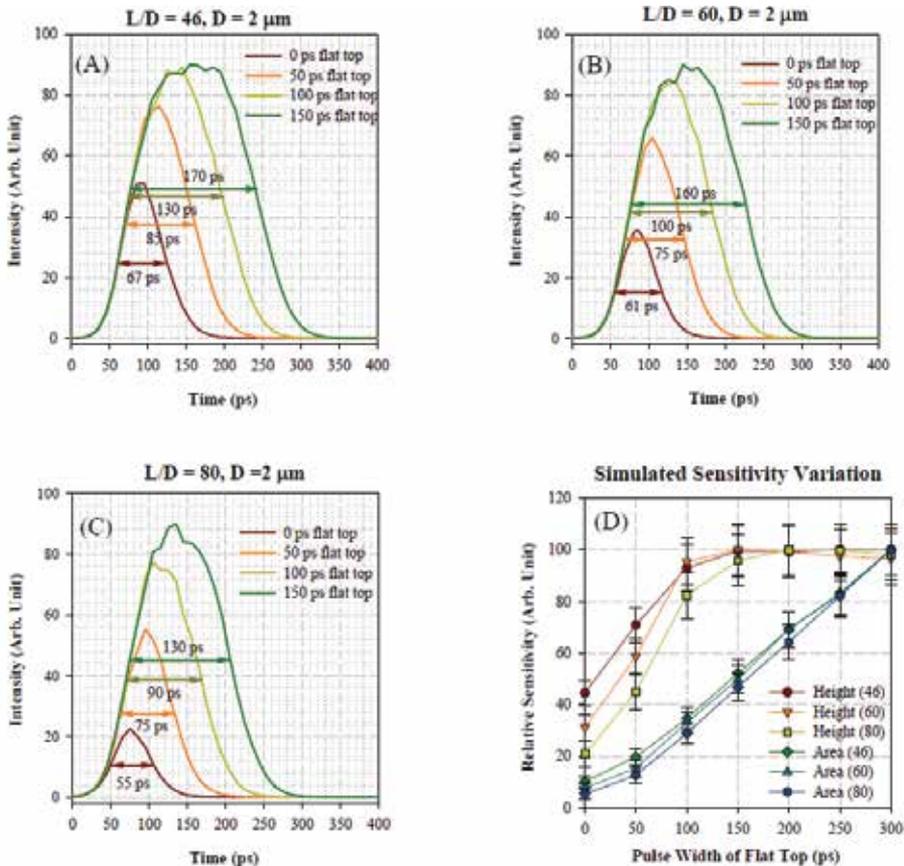


Fig. 17. Simulated optical gate profiles of  $2 \mu\text{m}$   $L/D = 46$ ,  $60$ , and  $80$  MCPs in (A), (B), and (C) using input voltage waveforms in Fig. 5(A); relative detection sensitivity in both peak height and integrated area for these MCPs are shown in (D).

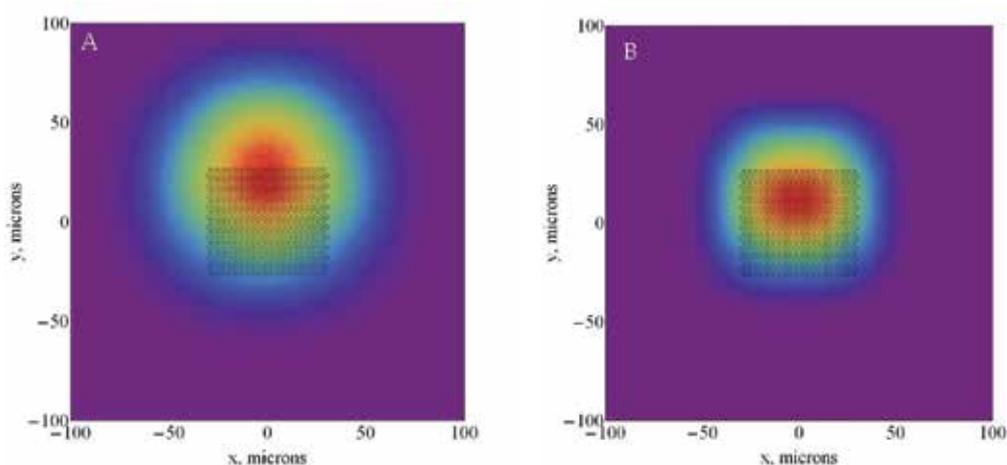


Fig. 18. Simulated output electron distributions on phosphor: (A)  $V_{ph} = 1000$  V; (B)  $V_{ph} = 4000$  V for  $2\ \mu\text{m}$   $L/D = 46$  MCP with a  $0.50$  mm gap between the MCP and phosphor.

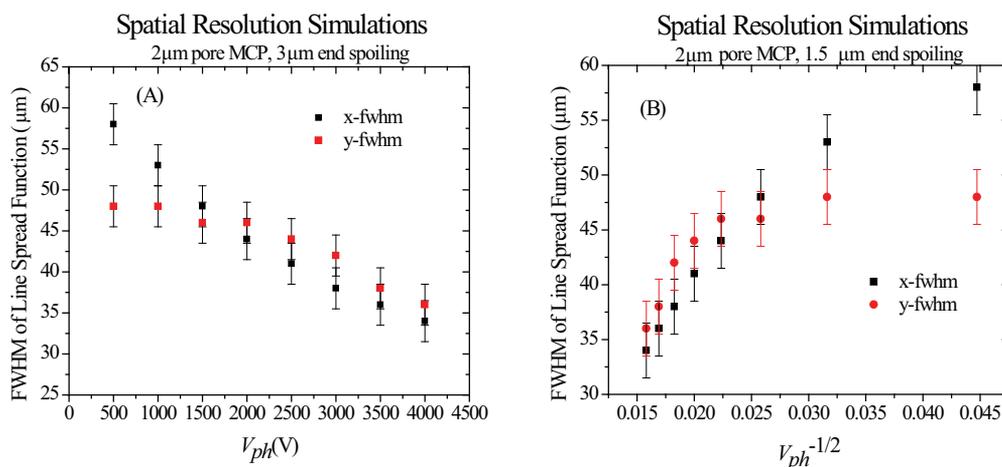


Fig. 19. Simulated spatial resolutions of  $2\ \mu\text{m}$   $L/D = 46$  MCP with  $3\ \mu\text{m}$  ( $1.5$  channel diameters) of end spoiling in parallel ( $x$ ) and perpendicular ( $y$ ) directions with bias angle of the pore and with a  $0.50$  mm gap between the MCP and phosphor (A)  $V_{ph}$  and (B)  $V_{ph}^{1/2}$ .

For the proximity focused approximation, one would expect the spatial resolution to increase linearly when the gap,  $d$ , between the MCP output and the phosphor is reduced (Kilkenny, 1991). For our final set of  $2\ \mu\text{m}$  pore MCP spatial resolution studies, the phosphor was placed closer to the MCP output, such that  $d = 0.25$  mm. This is indeed rather close and may not be physically achievable without arcing. However, the fact that the small-pore MCPs are usually rather small may make this MCP–phosphor distance obtainable. The results show that the LSF FWHM at  $V_{ph} = 2000$  V is near  $30\ \mu\text{m}$ , smaller than in the  $d = 0.5$  mm cases. This is a change of about  $15$  microns, somewhat less than the expected factor of two from the proximity focused approximation. The dependence versus  $V_{ph}^{-1/2}$  appears to be linear across the entire simulated range of  $V_{ph}$  in the  $x$ -direction, however.

In conclusion, it appears that small-pore MCPs may offer some significant improvements over 10  $\mu\text{m}$  pore MCPs for fast-gated imaging applications. In particular, much higher time resolutions may be achievable, with optical gate widths near 50 ps a possibility. Modest improvements on the order of 10  $\mu\text{m}$  in spatial resolution (measured as FWHM of the LSF) may also be achievable.

## 11. Conclusion

We have discussed our Monte Carlo simulation model, developed as a tool for assisting with the design of high-speed, time-gated x-ray cameras and for interpreting their data. The code uses a fairly standard set of equations for determining secondary emission yields, based in part on experimental data obtained with MCP lead glass. It also takes into account elastic reflections of low-energy electrons from the channel wall and requires that the total secondary electron energy not exceed the energy of the parent primary electron.

The plentiful experimental data we have for 0.46 mm thick, 10  $\mu\text{m}$  pore diameter MCPs allowed us to test our simulation code and fine-tune imprecisely known MCP secondary emission parameters. With this fine-tuning, our simulations of gain vs. DC bias voltage and of MCP spatial resolution achieve excellent agreement with experimental data.

In addition we studied the behavior of the electron cascade in an MCP under a time-dependent voltage pulse. We focused in particular on voltage pulses of duration shorter than or comparable to the DC electron transit time in an MCP. Simulations using ideal square waves of varying widths and with varying DC offsets illustrate some interesting behavior of MCPs in pulsed mode. In particular, we observe an increase in the MCP relative sensitivity (i.e., the MCP gain exponent) with peak voltage for pulse widths shorter than the transit time. However, this increase in the relative sensitivity is accompanied by a marked decrease in absolute sensitivity (the actual MCP gain). We see similar behavior in simulations using voltage pulses measured on MCP striplines. These latter simulations agree with experimental data obtained using a short-pulse UV laser.

Simulated and measured optical gate profiles for a gated x-ray camera with a 150 ps quasi-flat-top voltage pulse were presented. Voltage waveforms were measured at positions along an MCP stripline using a high-impedance, high-bandwidth probe. These waveforms were used in Monte Carlo code to calculate position-dependent gate profiles. Comparisons with gate profiles measured experimentally using a short-pulse UV laser demonstrated excellent agreement, both for the gate profile widths and the peak relative sensitivities.

We have also shown results of a simulation study of the potential performance of 2  $\mu\text{m}$  pore MCPs for fast-gated imaging applications. Results of 2  $\mu\text{m}$  pore MCP simulations were compared with simulated (and some measured) results from a 10  $\mu\text{m}$  pore MCP. We found that for the 2  $\mu\text{m}$  pore MCPs, gate profiles near 50 ps are potentially achievable with technology readily available. Additionally, we found that the simulated spatial resolutions are somewhat superior to what is achievable with a 10  $\mu\text{m}$  pore MCP. Thus, small-pore MCPs present an intriguing possibility for future HEDP imaging diagnostics.

## 12. Acknowledgments

This research has been partially supported by NSTec Nevada Test Site-Directed Research & Development (SDRD) funds and SNL Above-Ground Experimentation (AGEX) programs. The authors would like to thank Matt Griffin, Ken Moy, Shaun Hampton, and Andrew Mead for their assistance with the experimental measurements at the NSTec Livermore

Short Pulse Laser Facility, and Aric Tibbitts and Morris Kaufman for the H-CA-65 camera design. The authors would also like to thank Michele Vochosky, Robert Hilko, and Wilfred Lewis for critical reading of the manuscript.

This manuscript has been authored by National Security Technologies, LLC, under Contract No. DE-AC52-06NA25946 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

### 13. References

- Authinarayanan, A. & Dudding, R. (1976). Changes in secondary electron yield from reduced lead glasses. *Adv. Electron Phys. A*, 40, 167-181
- Bailey, J. E., Chander, G., Slutz, S., Golovkin, I., Lake, P., MacFarlane, R., Mancini, T., Burris-Mog, T., Cooper, G., Leeper, R., Mehlhorn, T., Moore, T., Nash, T., Nielsen, D., Ruiz, C., Schroen, D., & Varnum, W. (2004). Hot dense capsule-implosion cores produced by Z-pinch dynamic hohlraum radiation. *Phys. Rev. Lett.*, 92, 085002-1-4
- Bradley, D. K., Bell, P.M., Landen, O.L., Kilkenny, J.D., & Oertel, J. (1995). Development and characterization of a pair of 30 -40 ps x-ray framing cameras. *Rev. Sci. Instrum.* 66, 717-718
- Bruining, H. (1954). *Physics and Applications of Secondary Electron Emission*, Pergamon, 0080090141, London
- Bronshteyn, I., Yevdokimov, A., Stozharov, V., & Tyutikov, A. (1979). Differential secondary-emission characteristics of microchannel plates. *Radio Eng. Electron. Phys.*, 24, 150-152; 871-874
- Choi, Y. & Kim, J. (2000). Monte Carlo simulations for tilted-channel electron multipliers. *IEEE Trans. Elec. Dev.*, 47, 1293-1296
- Cimino, R., Collins, I., Furman, M., Pivi, M., Ruggiero, F., Rumolo, G., & Zimmerman, F. (2004). Can low energy electrons affect high energy physics accelerators? *Phys. Rev. Lett.*, 93, 014801-1-4
- Eberhardt, P. (1979). Gain model for microchannel plates. *Appl. Opt.*, 18, 1418-1423
- Fraser, G., Barstow, M., Whiteley, M., & Wells, A. (1982). Enhanced soft x-ray detection efficiencies for imaging microchannel plate detectors. *Nature (London)*, 300, 509-511
- Fraser, G., Barstow, M., Pearson, J., Whiteley, M., & Lewis, M. (1984). The soft x-ray detection efficiency of coated microchannel plates. *Nucl. Instrum. Methods*, 224, 272-286
- Gatti, E., Oba, K., & Rehak, P. (1983). Study of the electric field inside microchannel plate multipliers. *IEEE Trans. Nucl. Sci.*, NS-30, 461-468
- Guest, A. (1971). A computer model of channel multiplier plate performance. *Acta Electronica*, 14, 79-97
- Hirata, M., Cho, T., Takahashi, E., Yamaguchi, N., Kondoh, T., Matsuda, K., Aoki, S., Tanaka, K., Maezawa, H., & Miyoshi, S. (1992). X-ray detection characteristics of gold photocathodes and microchannel plates using synchrotron radiation (10 eV-82.5 keV). *Nucl. Instrum. Methods Phys. Res. B*, 66, 479-484
- Ito, M., Kume, H., & Oba, K. (1984). Computer analysis of the timing properties in micro channel plate photomultiplier tubes. *IEEE Trans. Nucl. Sci.*, NS-31, 408-412
- Kilkenny, J. (1991). High speed proximity focused X-ray cameras. *Laser and Particle Beams*, 9, 49-69

- Koshida, N. & Hosobuchi, M. (1985). Energy distribution of output electrons from a microchannel plate. *Rev. Sci. Instrum.*, 56, 1329-1331
- Koshida, N. (1986). Effects of electrode structure on output electron energy distribution of microchannel plates. *Rev. Sci. Instrum.*, 57, 354-358
- Kruschwitz, C. A., Wu, M., Moy, K., & Rochau, G. (2008). Monte Carlo simulations of high-speed, time-gated microchannel-plate based x-ray detectors: saturation effects in DC and pulsed modes and detector dynamic range. *Rev. Sci. Instrum.*, 79, 10E911-1 -4
- Landen, O., Bell, P., Oertel, J., Satariano, J., & Bradley, D. (1994). Gain uniformity, saturation and depletion in gated microchannel-plate x-ray framing cameras. *Proc. SPIE*, 2002, 2-13
- Landen, O., Abare, A., Hammel, B., Bell, P., & Bradley, D. (1993). Detailed measurements and shaping of gate profiles for microchannel-plate-based framing camera. *Proc. SPIE*, 2273, 245-254
- Landen, O., Lobban, A., Tutt, T., Bell, P., Costa, R., Hargrove, D., & Ze, F. (2001). Angular sensitivity of gated microchannel plate framing cameras. *Rev. Sci. Instrum.*, 72, 709 -
- McCarville, T., Fulkerson, S., Booth, R., Emig, J., Young, B., Anderson, S., & Heeter, B. (2005). Gated x-ray intensifier for large format simultaneous imaging. *Rev. Sci. Instrum.* 76, 103501-1 -6
- Melamid, A., Khachatryan, Z., Guzhov, A. (1972). Photoelectric emission in the wavelength region 500-1700 Å. *J. Appl. Spect.*, 16, 262-265
- Oertel, J., Aragonez, R., Archuleta, C., Barnes, C., Casper, L., Fatherley, V., Heinrichs, T., King, R., Landers, D., Lopez, F., Sanchez, P., Sandoval, G., Schrank, L., Walsh, P., Bell, P., Brown, M., Costa, R., Holder, J., Montelongo, S., & Pedersen, N. (2006). Gated x-ray detector for the National Ignition Facility. *Rev. Sci. Instrum.*, 77, 10E308-1-4
- Pawley, C. J & Deniz, D. V. (2000) Improved measurements of noise and resolution of x-ray framing cameras at 1 -2 keV. *Rev. Sci. Instrum.* 71. 1286-1296
- Price, G. & Fraser, G. (2001). Calculation of the output charge cloud from a microchannel plate. *Nucl. Instrum. Methods Phys. Res. A*, 474, 188-196
- Robey, H., Bundil, K., & Remington, B. (1997). Spatial resolution of gated x-ray pinhole cameras. *Rev. Sci. Instrum.*, 68, 792 -795
- Rochau, G., Bailey, J., Chandler, G., Nash, T., Nielsen, D., Dunham, G., Garcia, O., Joseph, N., Keister, J., Madlener, M., Morgan, D., Moy, K., & Wu, M. (2006). Energy dependent sensitivity of microchannel plate detectors. *Rev. Sci. Instrum.*, 77, 10E323-1-4
- Rochau, G., Wu, M., Kruschwitz, C., Joseph, N., Moy, K., Bailey, J., Crain, M., Thomas, R., Nielsen, D., & Tibbitts, A. (2008). Measurement and modeling of pulsed microchannel plate operation. *Rev. Sci. Instrum.*, 79, 10E902-1 - 6
- Scholtz, J., Dijkamp, D., & Schmitz, W. (1996). Secondary electron emission properties. *Philips. J. Res.*, 50, 375-389
- Shikhaliev, P. (1997). Hard x-ray detection model for microchannel plate detectors. *Nucl. Instrum. Meth. in Phys. Res. A*, 398, 229-237
- Vaughan, J. (1989). A new formula for secondary emission yield. *IEEE Trans. Elec. Devices*, 36, 1963-1967
- Wiza, J. (1979). Microchannel plate detectors. *Nucl. Instrum. Methods*, 162, 587 -601
- Wu, M., Kruschwitz, C., Morgan, D. , & Morgan, J. (2008). Monte Carlo simulations of MCP detectors. I. Steady-state voltage bias results. *Rev. Sci. Instrum.*, 79, 073104-1 -7
- Ze, F., Landen, O., Bell, P., Turner, R., Tutt, T., Alvarez, S., & Costa, R. (1999). Investigation of quantum efficiencies in multilayered photocathodes for microchannel plate applications. *Rev. Sci. Instrum.*, 70, 659-662

# Many-particle Monte Carlo Approach to Electron Transport

G. Albareda, F. L. Traversa, A. Benali and X. Oriols  
*Departament d'enginyeria Electrònica, Universitat Autònoma de Barcelona  
Spain*

## 1. Introduction

Recent technological advances have made possible to fabricate structures at the nanoscale. Such structures have already a large range of applications in very disparate fields of science and technology, and day by day new proposals are being suggested. A particular example of is nanoelectronics, where nanostructures constitute the platforms where smaller and smaller electron devices are being designed with the aim of letting the semiconductor industry to go forward in manufacturing faster and less consuming devices.

Today, due to the increase of the complexity and cost of the technological processes necessary to fabricate electron device prototypes, precise predictions on their functionality allowing to rule out specific designs are strictly necessary. In this regard, the success of nanoelectronics partially relies on physical theories on electron transport usually implemented on computer design tools that constitute at this moment a research and development cost reduction amount of the 35%, and is expected to increase up to 40% in the next future (Sverdlov et al., 2008). But more importantly, beyond the supporting role in the progress of electronics, theoretical approaches to electron transport constitute a necessary tool to guide the continuous breakthroughs of the electronics industry.

Analytical approaches to model electron transport have been developed since the invention of the first vacuum valve (Conwell, 1967), however, it has been the improvement of electronics itself which has make it possible to intensify the research on the simulation of electron transport by means of computer-aided tools. With the aid of faster computers, it became possible to obtain exact numerical solutions of complex microscopic physical models. The first fully numerical description of electron transport was already suggested in 1964 by Gummel (Gummel, 1964) for the one-dimensional bipolar transistor. The approach was further developed and applied to *pn*-junctions (De Mari, 1968) and to avalanche transit-time diodes by Scharfetter and Gummel (Scharfetter & Gummel, 1969). It was in 1966 that the first application of the Monte Carlo (MC) method was proposed by T. Kurosawa to solve the semi-classical Boltzmann Transport Equation (BTE) (Kurosawa, 1966). Since this pioneering work, the MC technique applied to device simulation has suffered a great evolution, and over more than 30 years it has been applied to understand several physical processes related with transport phenomena in many scenarios of interest (an extend review can be found in Refs. Jacoboni & Lugli, 1989; Jacoboni & Reggiani, 1983).

The use of the MC technique in the field of nanoelectronics is justified by the enormous complexity associated to the microscopical description of electron transport in solid-state structures. Current computational limitations forces any electron transport model to restrict the entire simulated degrees of freedom to a very few number. Such a simplification of the real scenario implies the use of probabilistic distributions that provide statistical information on those parts of the system that have been neglected/approximated<sup>1</sup>. Here is precisely where the MC technique comes into play by providing a sequence of random numbers with such particular distribution probabilities.

The study of electron transport through the BTE constitutes, however, just a single-particle approach to the electron transport problem (Albareda, Suñé & Oriols, 2009; Di Ventra, 2008), only accurate enough to describe electron dynamics in large structures containing a large number of carriers. As electron device sizes shrink into the nanometer scale, device structures are characterized by simultaneously holding a small number of electrons in a few nanometers. In this regime, carriers experience very little or no scattering at all (ballistic limit). The Coulomb interaction among carriers becomes then particularly important because the motion of one electron strongly depends on the motion of all the others and viceversa, i.e. their dynamics get strongly correlated. Hence, a many-particle formulation of electron transport becomes mandatory.

The main thrust of this chapter is to present a generalization of the standard MC solution of the BTE by introducing a rigorous many-particle description of electron dynamics and its classical correlations.

#### ***Overview on the treatment of Coulomb correlations:***

In the last decade, several works have remarked the necessity of paying maximum attention not only to electron-electron (e-e) but also to electron-impurity (e-i) Coulomb interactions when developing new approaches to electron transport (Barraud et al., 2002; Gross et al., 1999; 2000b; Wordelman & Ravaioli, 2000). In the MC solution of the BTE, historically the e-i (and also a part of the e-e) interactions have been introduced perturbatively as “instantaneous” and “local” transitions of electrons between different regions of the k-space. Such approach is clearly inappropriate at deep nanoscale because it assumes homogenous distributions of charges, so that the effects of “scattering” of electrons become position independent. The expressions for the e-e and e-i scattering rates in k-space are, in addition, based on a two-body model which accounts for many-particle contributions only through the screening function of simplified effective potentials, and which does not take into account the electrostatic effects of the gate, drain and source terminals on the potential distribution.

The standard solution to avoid such important limitations of the MC without significantly increasing the computational cost of the simulations, consists on defining an ad-hoc spatial division of the e-e and e-i interactions. For sufficiently large structures (several tens of nanometers), two distinguishable behaviors of the electrostatic potential can be identified depending on the proximity to the charges. On one hand, in a position very near ( $\sim$  few nanometers) to the carriers and ions, the shape of the scalar potential behaves just like  $1/r$ . On the other hand, far enough from them ( $\sim$  tens of nanometers), the shape of the scalar

---

<sup>1</sup> An example of these distributions is the *Fermi Dirac distribution*, which describes the way (i.e. position, time and momentum) electrons enter the active region once the simulation of the battery, leads and contacts has been avoided. Another example are, for instance, the *scattering rates* describing the interaction between electrons and the atoms conforming the underlying crystallographic mesh within the effective mass approximation.

potential depends on the particular spatial charge distribution. In this regard, a common strategy consists on introducing a long-range part of the Coulomb interaction numerically through the solution of the mesh-dependent mean-field Poisson equation, and a short-range part analytically (the so-called Monte Carlo/Molecular dynamics approach, MCMD, (Gross et al., 1999; Wordelman & Ravaioli, 2000)), or perturbatively (Barraud et al., 2002). The MCMD approach, however, shows some inconvenients. The first reported limitation was the so-called “double counting” of the electrostatic force in the short-range interaction term (Gross et al., 1999; 2000a). Since the e-e and e-i interactions are already included, in a smoothed way, in the self-consistent solution of the Poisson equation, the addition of a separate analytical force (the Molecular dynamics term) leads to the overestimation of both the e-e and the e-i interactions (Gross et al., 1999). Such a problem can be avoided by properly identifying the spatial region where short-range Coulomb interactions have to be included. In particular, the Molecular dynamics routine uses a “corrected” short-range Coulomb interaction that excludes the long-range contribution from the Poisson equation (Gross et al., 1999; 2000a;b). The problem comes then from the analytical nature of the short-range corrections, which can lead to unphysically large forces that cause artificial heating and cooling (for acceptors and donors respectively) of the carriers (Gross et al., 2000b; Ramey & Ferry, 2003). Again, this problem can be amended by introducing modifications of the analytical expressions of the Coulomb interaction in the short-range region (Alexander et al., 2005; 2008) or by implementing density gradient (quantum) corrections that accounts for the formation of bound states in the donor induced wells (Asenov et al., 2009; Vasileska & Ahmed, 2005).

Unfortunately, even all the above improvements of the MC solution of the BTE can fail when device dimensions are aggressively reduced to a very few nanometers either in lateral or longitudinal directions. Then, separations between long- (screened) and short- (unscreened) range contributions of the Coulomb interaction become quite misleading, and moreover, under such particular conditions, an important intrinsic limitation of the MC solution of the BTE come to light: it constitutes a single-particle description of electron transport, i.e. it describes the time-evolution of the electron distribution function (i.e. the Boltzmann distribution) in a single-electron phase-space (Albareda, López, Cartoixà, Suñé & Oriols, 2010; Albareda, Saura, Oriols & Suñé, 2010; Albareda, Suñé & Oriols, 2009; Boltzmann, 1872). Moreover, in addition to the above problems, due to the computational burden associated to the microscopic description of electron transport, the simulation of the Coulomb correlations between the electrons in the leads and those in the active region of an electron device is not always possible and the use of small simulation boxes is a mandatory requirement in modern nanoscale simulators (see Fig. 1). However, in order to correctly model the DC and/or AC conductance of nanoscale systems, one has to assure the accomplishment of “overall charge neutrality” and “current conservation” (Blanter & Büttiker, 2000; Landauer, 1992). The implementation of such requirements into nanoscale electron simulators demands some kind of reasonable description of the Coulomb interaction among the electrons inside and outside the simulation boxes. The boundary conditions on the borders of simulation boxes in electron transport approaches constitute also a complicate and active field of research. In fact, educated guesses for the boundary conditions are present in the literature when describing nanoscale electron devices with simulation boxes large enough to include the leads. However, such boundary conditions are not applicable for small simulation boxes that exclude the leads. Elaborated semi-classical electron transport simulators solving the time-dependent BTE within the MC technique commonly fix the potential at the borders of the simulation box equal to the external bias (i.e. Dirichlet boundary conditions) and assume ad-hoc modifications of

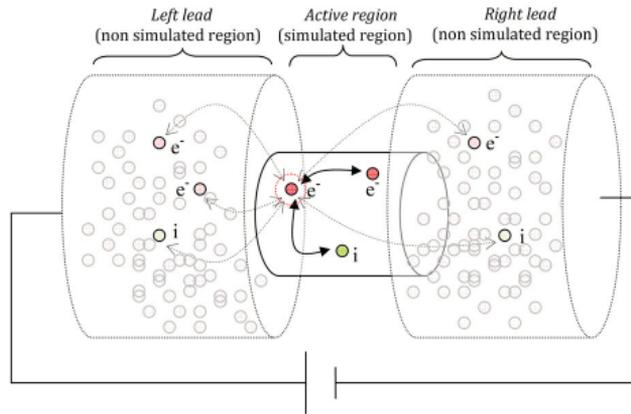


Fig. 1. Schematic representation of the Coulomb correlations among electrons in the leads and those in the active region of an electron device.  $e^-$  and  $i$  represent the conduction electrons and the ionized atoms respectively.

the injection rate to achieve “local” charge neutrality (Bulashenko et al., 1998; Fischetti & Laux, 2001; Gomila et al., 2002; Gonzalez et al., 1997; Gonzalez & Pardo, 1993; 1996; Jacoboni & Lugli, 1989; Reklaitis & Reggiani, 1999; Wordelman & Ravaioli, 2000). Some works do also include analytically the series resistances of a large reservoir which can be considered an improvement over the previous boundary conditions (Babiker et al., 1996). Other MC simulators do also consider Neumann boundary conditions (i.e. a fixed zero electric-field) (Riddet et al., 2008). The latter conditions fix also the scalar potential (up to an arbitrary constant) so that the injected charge can also be indirectly determined when a known electrochemical potential is assumed. Although all these boundary conditions are successful for large simulation boxes, they are quite inaccurate for small simulation boxes that exclude the leads (Riddet et al., 2008). In principle, there are no much computational difficulties in applying a semi-classical MC technique in large simulation boxes when dealing with mean-field approaches. However, the possibility of using smaller boxes will be very welcomed for some intensive time-consuming simulations beyond the mean-field approximation (also for statistical ensemble simulations, see Ref. Reid et al., 2009, or to compute current or voltage fluctuations that need very large simulation times to obtain reasonable estimators, see Refs. Albareda, Jiménez & Oriols, 2009; Gonzalez et al., 1997; Gonzalez & Pardo, 1993; Reklaitis & Reggiani, 1999; Wordelman & Ravaioli, 2000, etc.).

In this chapter, we are interested in revisiting the MC computation of an ensemble of Coulomb interacting particles in an open system (an electron device) without any of the approximations mentioned in the previous paragraphs. With this goal, in section 2, we will develop an exact many-particle Hamiltonian for Coulomb interacting electrons in open systems in terms of the solutions of multiple Poisson equations (Albareda, Suñé & Oriols, 2009). To our knowledge, the type of development of the many-particle Hamiltonian proposed here has not been previously considered in the literature because, up to now, it was impossible to handle the computational burden associated with a direct solution of a many-particle Hamiltonian. Furthermore, in section 3, we will present an original (time-dependent) boundary condition algorithm for open systems capable of accurately capturing Coulomb correlations among the electrons inside and outside the simulation box (Albareda, López, Cartoixà, Suñé & Oriols,

2010). Such boundary conditions constitute a notable improvement of standard boundary conditions used in MC approaches and, requiring a minimum computational effort, they can be implemented into time-dependent simulators with large or small simulation boxes, and for DC, AC conditions and even for the study of current (or voltage) fluctuations. Section 4, will be devoted to present a semi-classical solution of the time-dependent many-particle Hamiltonian provided with the previous boundary conditions. This solution constitutes a generalization of the semi-classical single-particle Boltzmann distribution to many-particle systems (Albareda, López, Cartoixà, Suñé & Oriols, 2010; Albareda, Suñé & Oriols, 2009). In section 5, our many-particle MC will be used to evaluate the importance of accounting for strongly-correlated phenomena when predicting discrete doping induced effects in the channel of a quantum wire double-gate field-effect transistor (Albareda, Saura, Oriols & Suñé, 2010).

## 2. A Many-particle Approach to Semi-Classical Electron Transport

Although interacting many-electron systems are well assessed through the exact expression of the system's Hamiltonian, its exact solution is a very hard problem when the number of interacting electrons increase farther than a few tens. Therefore, the main difficulty that one encounters when describing electron transport at the nanoscale arises from the necessity of making reasonable approximations to an essentially untractable problem, i.e. the many-body description of electron transport.

Consider, for instance, the Hamiltonian describing a whole closed circuit, i.e. including the battery, the contacts, the leads, the active region and all the constituting elements therein (see Fig. 1). If we assume that it contains  $M_T$  electrons and  $W - M_T$  atomic cores, the Hamiltonian of the system can be written as

$$\begin{aligned}
 H_{circuit}(\vec{r}_1, \dots, \vec{r}_G, \vec{p}_1, \dots, \vec{p}_G) &= \sum_{k=1}^{M_T} \left\{ K(\vec{p}_k) + \frac{1}{2} \sum_{\substack{j=1 \\ j \neq k}}^{M_T} eV_0(\vec{r}_k, \vec{r}_j) \right\} \\
 &+ \sum_{k=M_T+1}^W \left\{ K(\vec{p}_k) + \frac{1}{2} \sum_{\substack{j=M_T+1 \\ j \neq k}}^W eZ_k Z_j V_0(\vec{r}_k, \vec{r}_j) \right\} \\
 &- \sum_{k=1}^{M_T} \sum_{j=M_T+1}^W eZ_j V_0(\vec{r}_k, \vec{r}_j), \tag{1}
 \end{aligned}$$

where  $K(\vec{p}_k)$  is the kinetic energy of the  $k$ -th particle with a momentum  $\vec{p}_k$ ,  $e$  is the electron charge,  $\vec{r}_k$  is the vector position of the  $k$ -th particle, and  $Z_k$  is the atomic number of the  $k$ -th atom. The term  $V_0(\vec{r}_k, \vec{r}_j) = e / (4 \pi \epsilon_0 |\vec{r}_k - \vec{r}_j|)$  is the Coulomb potential (with  $\epsilon_0$  the vacuum permittivity)<sup>2</sup>.

<sup>2</sup> Along the whole chapter we will assume that all involved electrons are traveling at velocities much lower than light's,  $c$  (nonrelativistic approximation). Moreover, we will consider that we can neglect the electron spin-orbit coupling and that a quasi-static electromagnetic regime can be assumed. Such an assumption, however, does not mean that we are considering spinless electrons. Indeed, when computing some relevant magnitudes we will account for the electron spin just by an additional factor 2. (see Ref. Albareda, López, Cartoixà, Suñé & Oriols, 2010)

## 2.1 The many-electron open system Hamiltonian

Notice that the solution of the Hamiltonian (1) constitute an insurmountable challenge. The reason, however, does not only reside on the huge number of variables conforming the system ( $W \rightarrow \infty$ ), but mainly on their correlation, which does not allow to separate the dynamics of electrons. The interaction terms in (1) are the responsible of coupling each particle dynamics to the rest of particle dynamics in the system, and hence the responsible of making their solution so difficult. A common strategy to simplify the solution of equation (1) consists on reducing the involved degrees of freedom. Instead of trying to describe a whole *closed* circuit, *all* approaches to electron transport reduce the number of variables to be explicitly described by *opening* the system. As a first approach, we delimit the particles dynamics to be explicitly described to those that are free to carry electrical current.

**Internally opening the system:** In order to reduce the degrees of freedom in (1), we first remove their explicit dependence on the valence and core electrons by modifying the vacuum permittivity ( $\epsilon_0 \rightarrow \epsilon = \epsilon_r \cdot \epsilon_0$ , where  $\epsilon_r$  is the relative permittivity) and accounting for an average induced polarization between the bounded electrons and the nuclei (Reitz et al., 1992)). We also assume the adiabatic approximation (Datta, 1995; 2005; Lundstrom & Guo, 2006) (also called Born-Oppenheimer approximation) under which conducting electrons move in a quasi-static atomic potential defined by the fixed positions of the atoms. The original Hamiltonian (1), has been then reduced to the Hamiltonian of the carriers alone:

$$H_{carriers}(\vec{r}_1, \dots, \vec{r}_M, \vec{p}_1, \dots, \vec{p}_M) = \sum_{i=1}^M \left\{ K(\vec{p}_i) + \frac{1}{2} \sum_{\substack{j=1 \\ j \neq i}}^M eV(\vec{r}_i, \vec{r}_j) - \sum_{j=M_T+1}^W eZ_j V(\vec{r}_i, \vec{R}_j) \right\} \quad (2)$$

where  $M$  is now the total number of unbounded electrons, and  $R_j$  are now the fixed positions of the atoms. The Coulomb potential,

$$V(\vec{r}_i, \vec{r}_j) = \frac{e}{4\pi\epsilon |\vec{r}_i - \vec{r}_j|} \quad (3)$$

has been properly redefined according to the effective value of the dielectric permittivity. From now on, the dynamics of the nucleus and the bounded electrons are not anymore explicitly accounted for. Therefore, we do not deal with the circuit Hamiltonian,  $H_{circuit}$ , but with that of the  $M$  unbounded electrons (i.e. carriers),  $H_{carriers}$ . Although the spatial region described by  $H_{carriers}$  is the same as the one described by  $H_{circuit}$ , we have substantially reduced its complexity by disregarding many *internal* degrees of freedom. Hence the title "internally opening the system".

We are however still dealing with a computationally insolvable problem. In order to continue reducing the degrees of freedom, an important decision to be made is that of the energy band model that will be used. Assuming the electron-atom interaction potential to be an average over a unit cell of the atomic lattice of the semiconductor, carrier kinetics can be treated almost in the same way as for free carriers, but with a modified mass called the effective mass. Such a model is thus often called the effective mass model. We define then  $H_0$  as that part of the whole Hamiltonian containing the kinetic terms and the interaction among electrons and

atoms. We can rewrite (2) as

$$\hat{H}_{carriers} = \hat{H}_0 + \frac{1}{2} \sum_{i=1}^M \sum_{\substack{j=1 \\ j \neq i}}^M eV(\vec{r}_i, \vec{r}_j), \quad (4)$$

where

$$\hat{H}_0 = \sum_{i=1}^M H_{0i} = \sum_{i=1}^M \left\{ K(\vec{p}_i) - \sum_{j=M_T+1}^G eZ_j V(\vec{r}_i, \vec{R}_j) \right\}. \quad (5)$$

$H_0$  is separable and we can find mono-electronic eigenstates for every one of the  $M$  Hamiltonians  $H_{0i}$ . Moreover, if we assume an ideal periodic atomic structure, the Bloch states are solutions of these mono-electronic Hamiltonians. Considering that only one single-band is attainable by the carriers, it can be shown that the Hamiltonian (2) can be reduced to that of a many-particle envelope function (Albareda, 2010)

$$\hat{H}_{env} = \sum_{j=1}^M \left[ K(\vec{p}_k) + \frac{1}{2} \sum_{i=1}^M eV(\vec{r}_i, \vec{r}_j) \right], \quad (6)$$

where  $K(\vec{p}_k)$  is redefined as

$$K(\vec{p}_k) = -\frac{\hbar^2}{2} \left( \frac{\partial^2}{m_x^* \partial x_j^2} + \frac{\partial^2}{m_y^* \partial y_j^2} + \frac{\partial^2}{m_z^* \partial z_j^2} \right), \quad (7)$$

and  $m_x$ ,  $m_y$  and  $m_z$  are the trace terms of the diagonalized effective mass tensor (Albareda, 2010).

The Hamiltonian in (6) is still computationally unaffordable because it involves a huge number of degrees of freedom (those of the battery, contacts, leads, etc...). In this regard, we must carry out a step forward in simplifying the described electronic system. In what follows, we spatially delimit the system to be described, i.e. we externally open the electronic system.

**Externally opening the system:** We divide the previous ensemble of  $M$  particles into a sub-ensemble of  $N(t)$  particles whose positions are inside the volume  $\Omega$  and a second sub-ensemble,  $\{N(t) + 1, \dots, M\}$  which are outside (see figure 2). We assume that the number of particles inside,  $N(t)$ , is a time-dependent function that provides an explicit time-dependence to the many-particle (open-system) Hamiltonian. As drawn in figure 2, we define a parallelepiped where the six rectangular surfaces  $S = \{S^1, S^2, \dots, S^6\}$  are the boundaries of  $\Omega$ . We use  $\vec{r}^l$  as the ‘‘boundary’’ vector representing an arbitrary position on the surfaces  $S^l$ . Now, the number of carriers in the system  $N(t)$  will vary with time, i.e.

$$\hat{H}_{env}^{open}(\vec{r}_1, \dots, \vec{r}_M, \vec{p}_1, \dots, \vec{p}_{N(t)}, t) = \sum_k \left\{ K(\vec{p}_k) + \frac{1}{2} \sum_{\substack{k=1 \\ k \neq j}}^{N(t)} eV(\vec{r}_k, \vec{r}_j) + \sum_{j=N(t)+1}^M eV(\vec{r}_k, \vec{r}_j) \right\} \quad (8)$$

As we will be discussed below, the third term in (8) can be included in the Hamiltonian of the open system through the boundary conditions of the Poisson equation. The roughness of the approximation bringing together the effects of all the external particles over the  $N(t)$  carriers

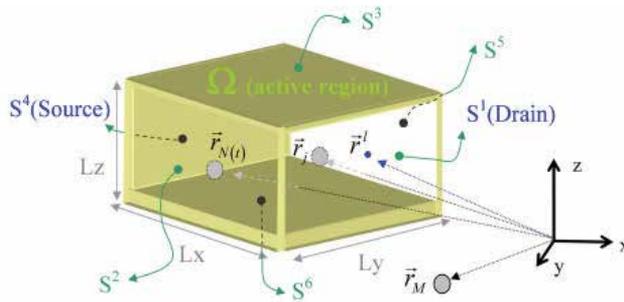


Fig. 2. Schematic representation of the open volume  $\Omega = L_x \cdot L_y \cdot L_z$  and its limiting surface  $S = \{S^1, S^2, \dots, S^6\}$ . There are  $N(t)$  particles inside and  $M - N(t)$  outside this volume. The vector  $\vec{r}^l$  points to an arbitrary position at the boundary surface  $S^l$ . We consider  $S^4$  and  $S^1$  as “opened” surfaces because there is a flux of particles through them. The rest of the surfaces are “closed” borders.

will depend on our ability of formulating the boundary conditions at the borders of the active region of the electron device (see Section 3 for an extent discussion on this point). Since we will only work with the many-particle open system Hamiltonian (8), in order to simplify the notation let us simply refer to it as  $\hat{H}$ .

In the previous paragraphs, the assumption of a series of approximations have make it possible to go from the complex circuit Hamiltonian described in equation (1), to the much more simple one describing the “interesting” region of the circuit (8). However, although we have discussed the many-particle Hamiltonian in terms of the Coulomb force, this approach is inconvenient to deal with solid-state scenarios with a spatial-dependent permittivity (Javid & Brown, 1963). For this reason, in what follows we rewrite the many-particle Hamiltonian (8) in terms of the more generic Poisson equation, which can be applied to systems with a spatial-dependent permittivity (by simply substituting  $\epsilon \rightarrow \epsilon(\vec{r})$ ). We start by rewriting the previous many-particle open system Hamiltonian (8) as:

$$H(\vec{r}_1, \dots, \vec{r}_M, \vec{p}_1, \dots, \vec{p}_{N(t)}, t) = \sum_{k=1}^{N(t)} \left\{ K(\vec{p}_k) + \sum_{\substack{j=1 \\ j \neq k}}^{N(t)} e \cdot V(\vec{r}_k, \vec{r}_j) + \sum_{j=N(t)+1}^M e \cdot V(\vec{r}_k, \vec{r}_j) - \frac{1}{2} \sum_{\substack{j=1 \\ j \neq k}}^{N(t)} e \cdot V(\vec{r}_k, \vec{r}_j) \right\}. \quad (9)$$

Each term  $V(\vec{r}_k, \vec{r}_j)$  that appears in (9) can be explicitly obtained from a Poisson (or Laplace) equation inside the volume  $\Omega$ . Using the superposition property of the Poisson equation, we can rewrite (9) as:

$$H(\vec{r}_1, \dots, \vec{r}_{N(t)}, \vec{p}_1, \dots, \vec{p}_{N(t)}, t) = \sum_{k=1}^{N(t)} \left\{ K(\vec{p}_k) + e \cdot W_k(\vec{r}_1, \dots, \vec{r}_{N(t)}, t) - \frac{1}{2} \sum_{\substack{j=1 \\ j \neq k}}^{N(t)} e \cdot V(\vec{r}_k, \vec{r}_j) \right\}, \quad (10)$$

where the term  $W_k(\vec{r}_1, \dots, \vec{r}_k, \dots, \vec{r}_{N(t)})$  is a particular solution of the following Poisson equation:

$$\nabla_{\vec{r}_k}^2 \left( \varepsilon \cdot W_k(\vec{r}_1, \dots, \vec{r}_{N(t)}) \right) = \rho_k(\vec{r}_1, \dots, \vec{r}_{N(t)}) . \quad (11)$$

The term  $\rho_k(\vec{r}_1, \dots, \vec{r}_{N(t)})$  in (11) depends on the position of the first  $N(t)$  electrons, i.e.

$$\rho_k(\vec{r}_1, \dots, \vec{r}_k, \dots, \vec{r}_{N(t)}) = \sum_{\substack{j=1 \\ j \neq k}}^{N(t)} e \cdot \delta(\vec{r}_k - \vec{r}_j), \quad (12)$$

however (12) is independent of the position of the external particles because they only affect the boundary conditions of (11). Let us notice that there are still terms,  $V(\vec{r}_k, \vec{r}_j)$ , in (10) that are not computed from the Poisson equations in (11), but from (3). If we compare expressions (9) and (10), the term  $W_k(\vec{r}_1, \dots, \vec{r}_{N(t)}, t)$  can be rewritten as:

$$W_k(\vec{r}_1, \dots, \vec{r}_{N(t)}, t) = \sum_{\substack{j=1 \\ j \neq k}}^{N(t)} V(\vec{r}_k, \vec{r}_j) + \sum_{i=N(t)+1}^M V(\vec{r}_k, \vec{r}_i) \quad (13)$$

The dependence of  $W_k(\vec{r}_1, \dots, \vec{r}_{N(t)}, t)$  on the positions of the external particles is explicitly written in the last sum in (13), while in (11) this dependence is hidden in the boundary conditions of  $W_k(\vec{r}_1, \dots, \vec{r}_k, \dots, \vec{r}_{N(t)})$  on the surface  $S = \{S^1, S^2, \dots, S^6\}$ . In fact, the boundary conditions are a delicate issue that we will discuss in the next section 3.

## 2.2 Comparison between the many-particle and the standard single-particle Monte Carlo

In this sub-section we emphasize the important differences appearing between the standard MC solution of the BTE, i.e. a time-dependent mean-field algorithm, and our time-dependent many-particle algorithm, i.e. a generalization of the previous single-particle formulation. With this aim in mind, let us first introduce the mean-field version of expression (10).

As described in the introduction of this chapter, the mean-field approximation provides a single average potential for computing the dynamics of all electrons. This average potential, that we label here with the suffix "mean",  $\bar{W}_{mean}(\vec{r}, t)$ , is still capable of preserving most of the collective effects of the Coulomb interaction. The term  $\bar{W}_{mean}(\vec{r}, t)$  is computed by taking into account all charges inside the volume  $\Omega$ . However, since one particle can not "feel" its own charge,  $\bar{W}_{mean}(\vec{r}, t)$  can be interpreted as the electrostatic potential "seen" by an additional probe charge whose position is  $\vec{r}$ , i.e.

$$\bar{W}_{mean}(\vec{r}, t) = \bar{W}_{M+1}(\vec{r}_1[t], \dots, \vec{r}_{N(t)}[t], \vec{r}), \quad (14)$$

where  $\bar{W}_{mean}(\vec{r}, t)$  is a solution of a unique 3D-Poisson equation:

$$\nabla_{\vec{r}}^2 \bar{W}_{mean}(\vec{r}, t) = \bar{\rho}_{mean}(\vec{r}, t), \quad (15)$$

and the charge density is defined as:

$$\bar{\rho}_{mean}(\vec{r}, t) = \sum_{j=1}^{N(t)} q_j \delta(\vec{r} - \vec{r}_j[t]), \quad (16)$$

Now, it can be shown that the error of the time-dependent mean-field approximation,  $Error_k(\vec{r}, t) = \bar{W}_{mean}(\vec{r}, t) - W_k(\vec{r}, t)$ , is (Albareda, Suñé & Oriols, 2009):

$$Error_k(\vec{r}, t) = \frac{1}{N(t)} \sum_{j=1}^{N(t)} \left\{ \left( \bar{W}_j(\vec{r}, t) - \bar{W}_k(\vec{r}, t) \right) + V(\vec{r}, \vec{r}_j[t]) \right\} = V(\vec{r}, \vec{r}_k[t]), \quad (17)$$

Expression (17) shows that  $Error_k(\vec{r}, t) \rightarrow \infty$ , when  $\vec{r} \rightarrow \vec{r}_k[t]$ . The above mean-field approximation implies that the potential “felt” by the  $k$ -particle at  $\vec{r} \rightarrow \vec{r}_k[t]$  is its own potential profile. In fact, from a numerical point of view, the use of the mean-field approximation is not so bad. For example, classical simulators uses 3D meshes with cell sizes of a few nanometers,  $DX \approx DY \approx DZ \gg 10 \text{ nm}$ . Then, the error of the mean-field approximation is smaller than the technical error (i.e. mesh error) due to the finite size of the cells. The long range Coulomb interaction is well captured with the mean-field approximation, while this approximation is a really bad strategy to capture the short range Coulomb interaction. Finally, let us remark another important point about the mean-field approximation. Looking to the final expression (17), rewritten here as  $W_k(\vec{r}, t) = \bar{W}_{mean}(\vec{r}, t) - V(\vec{r}, \vec{r}_k[t])$ , it seems that  $W_k(\vec{r}, t)$  can be computed from a unique mean-field solution of the Poisson equation  $\bar{W}_{mean}(\vec{r}, t)$  when subtracting the appropriate two-particle potential  $V(\vec{r}, \vec{r}_k[t])$ . However, such a strategy is not as general as our procedure because it requires an analytical expression for the two-particle Coulomb interaction  $V(\vec{r}, \vec{r}_k[t])$ . The analytical expression of  $V(\vec{r}, \vec{r}_k[t])$  written in expression (3) is only valid for scenarios with homogenous permittivity. On the contrary, our procedure with  $N(t)$  electrostatic potentials computed from  $N(t)$  different Poisson equations in a limited 3D volume  $\Omega$  can be applied inside general scenarios with spatial dependent permittivity.

### 2.2.1 Simulation of a two-electron system

In order to clarify the above discussion, let us consider one electron (labeled as 1-electron) injected from the source surface,  $S_4$ , at an arbitrary position (see Fig. 2). A second electron is injected from the drain surface,  $S_1$  (see Fig. 2). A battery provides an external voltage equal to zero at the drain and source surface. A 3D cubic system with a volume of  $\Omega = (20 \text{ nm})^3$  is considered as the active device region. We consider Silicon parameters for the numerical simulation. Within the mean-field approximation only the potential profile  $\bar{W}_{mean}(\vec{r}, t)$  is calculated for the two-electron system using expressions (14)-(16). Then, we realize from figure 3 that each electron can be reflected by an artificial alteration of the potential profile related to its own charge. In figures 4 and 5 we have plotted the energy potential profile “seen” by the 1-electron,  $\bar{W}_1(\vec{r}_1, t)$ , and by the 2-electron,  $\bar{W}_2(\vec{r}_2, t)$ , using the many-particle algorithm described in expressions (25)-(28). Electrons are not longer affected by their own charge. We clearly see that, within the mean-field approximation, electrons could even be unable to overcome the large potential barrier that appears at their own position (due to their own charge). In addition, these simple results confirm that the mean-field error is equal to expression (17), i.e. the error of the mean-field potential profile at each position of the active region is  $Error_k(\vec{r}, t) = V(\vec{r}, \vec{r}_k[t])$ .

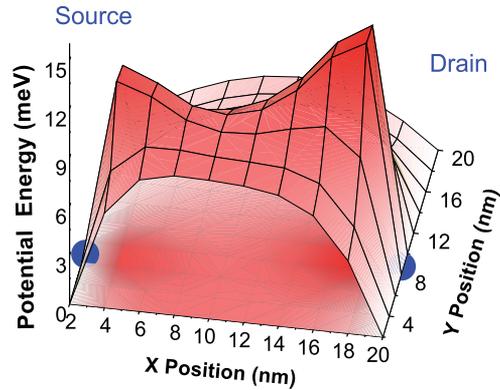


Fig. 3. Potential energy profile  $\bar{W}_{mean}(\vec{r}, t)$  computed with a 3D Poisson solver using the classical “mean-field” approximation on the plane X-Y of the active region  $\Omega = (20 \text{ nm})^3$  at  $z=6\text{nm}$  at 0.4 fs. The solid points are electron positions. Reprinted with permission from G. Albareda et al., Phys. Rev. B. 79, 075315 (2009). ©Copyright 2009, American Physical Society.

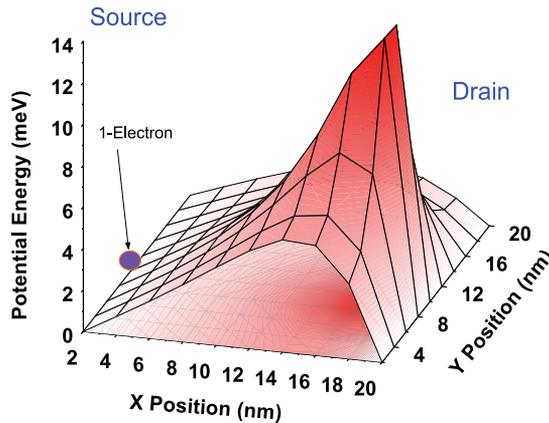


Fig. 4. Potential energy profile of the 1-electron,  $\bar{W}_1(\vec{r}_1, t)$ , with the “many-electron” algorithm in the plane X-Y of the active region  $\Omega = (20 \text{ nm})^3$  at  $z=6\text{nm}$  at 0.4 fs. The solid point is the 1-electron position. Reprinted with permission from G. Albareda et al., Phys. Rev. B. 79, 075315 (2009). ©Copyright 2009, American Physical Society.

Finally, a discussion about the role of the spatial mesh used for the numerical solution of the Poisson equation is relevant. For an electron device with a length of hundreds of nanometers, we need a mesh of the 3D active region with spatial step  $DX \sim DY \sim DZ > 10 \text{ nm}$  to deal with no more than one thousand nodes in the numerical solution of the Poisson equation. This computational limitation in the resolution of the potential is present either when solving the mean-field or the many-electron algorithm. With such spatial resolution, the short-range interaction is missing in both procedures because two electrons inside the same spatial cell will

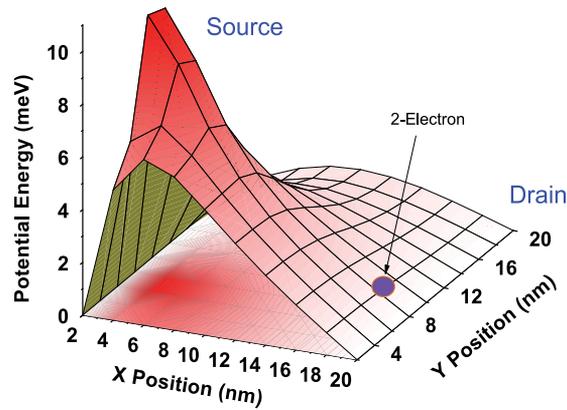


Fig. 5. Potential energy profile of the 2-electron,  $\bar{W}_2(\vec{r}_2, t)$ , with the “many-electron” algorithm in the plane X-Y of the active region  $\Omega = (20 \text{ nm})^3$  at  $z=6\text{nm}$  at  $0.4 \text{ fs}$ . The solid point is the 1-electron position. Reprinted with permission from G. Albareda et al., Phys. Rev. B. 79, 075315 (2009). ©Copyright 2009, American Physical Society.

not repel each other. In addition, the error between both procedures,  $Error_k(\vec{r}, t) = V(\vec{r}, \vec{r}[t]_k)$ , is reduced because the numerical Coulomb potential profile is smoothed due to the low resolution (i.e. the diameter of the region where  $V(\vec{r}, \vec{r}[t]_k)$  has a strong influence is shorter than the cell dimensions). Therefore, we obtain roughly identical results with both schemes. On the other hand, for better mesh resolutions ( $DX = DY = DZ = 2 \text{ nm}$ ) associated to smaller devices, the differences between both treatments increase due to the important spurious auto-reflection effect found in the mean-field trajectory. In summary, when the spatial cells are large, the mean-field and the many-electron schemes correctly model the long-range Coulomb interaction, but both neglect the short-range component. On the contrary, with smaller spatial steps  $DX \sim DY \sim DZ < 5 \text{ nm}$ , the many-electron resolution takes into account long- and short- range Coulomb interaction correctly, whereas the description of the short-range component within the mean-field approximation is completely incorrect (i.e. electrons are repelled by themselves). In other words, when  $DX, DY, DZ \rightarrow 0$  the mesh error in our many-electron algorithm reduces to zero, while the error in the mean-field approach tends to infinite,  $Error_k(\vec{r}, t) \rightarrow \infty$ .

### 3. Boundary conditions for the many-particle open system Hamiltonian

Let us now move back to the expressions defining our transport problem, i.e. (10), (11) and (12). In order to self-consistently solve electron dynamics in our open system, we need the solutions of the Poisson equations (11). In this regard, the boundary conditions of the  $N(t)$  terms  $W_k(\vec{r}_1, \dots, \vec{r}_k, \dots, \vec{r}_{N(t)})$  in (13) must be specified on the border surfaces  $S = \{S^1, S^2, \dots, S^6\}$  of figure 2. Such boundary conditions will provide valuable information on the electrostatic effect that the electrons plus the impurities outside have on the electrons inside  $\Omega$ .

In practical situations, the volume  $\Omega$  describes the active region of some kind of electron device (a MOSFET for instance). In order to discuss our boundary conditions algorithm, we

assume here a two-terminal device (source and drain)<sup>3</sup>. This means that only two,  $S^1$  and  $S^4$ , of the six border surfaces  $S = \{S^1, S^2, \dots, S^6\}$  are really opened to the flow of carriers (see figure 6). These opened surfaces represent a complicate problem of boundary conditions that will be discussed in detail in this section (see also Ref. Albareda, López, Cartoixà, Suñé & Oriols, 2010)<sup>4</sup>.

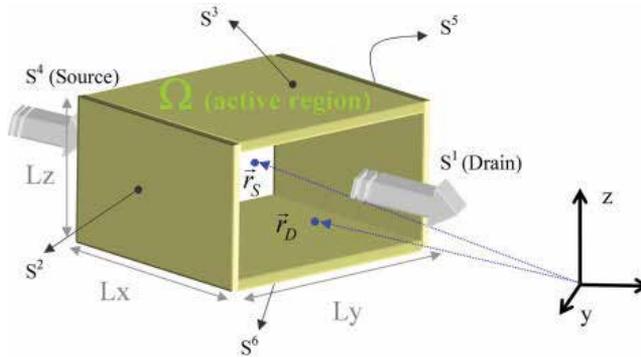


Fig. 6. Schematic representation of the volume  $\Omega = L_x \cdot L_y \cdot L_z$  representing a two terminal device. Only  $S^1$  and  $S^4$ , corresponding to the drain and source surfaces respectively, are opened to electron flow. On the rest of surfaces standard Neumann boundary conditions are assumed.

### 3.1 On the importance of boundary conditions

In order to correctly model the DC and/or AC conductance of nanoscale systems, one has to assure the accomplishment of “overall charge neutrality” and “current conservation” (Blanter & Büttiker, 2000; Landauer, 1992). As we have mentioned, the implementation of such requirements into modern nanoscale electron simulators demands some kind of reasonable approximation for the Coulomb interaction.

In general, modern electron transport simulators do include reasonable approximations for the coulomb interactions that can guarantee the accomplishment of the “overall charge neutrality” requirement. In addition, those simulators that are developed within a time-dependent or frequency-dependent framework can also assure the “current conservation” requirement. However, the treatment of many-particle electron transport can only be applied to a very limited number of degrees of freedom. In fact, due to computational restrictions, a small simulation box is a mandatory requirement in modern MC simulators. This restriction implies that either very short leads (with screening length of few Armstrongs) are included into the small simulation box, or the leads are directly excluded from the simulation box. The first solution is only acceptable for metallic leads (Brandbyge et al., 2002;

<sup>3</sup> In any case, the boundary conditions can be straightforwardly adapted to multi-terminal systems with an arbitrary number of “opened” borders.

<sup>4</sup> On the “closed” non-metallic surfaces<sup>5</sup>, Neumann boundary conditions are used with the educated guess that the component of the electric field normal to that surfaces is zero. The continuity of the displacement vector normal to surfaces justifies this assumption on “closed” (i.e. no electrons traversing the surfaces) boundaries when the relative permittivity inside is much higher than the corresponding value outside. On “open” metallic surfaces, we use a many-particle version of the standard Dirichlet boundary conditions (Albareda, Suñé & Oriols, 2009)

Taylor et al., 2001) close to equilibrium, but it becomes inappropriate in general scenarios ranging from highly doped poly-silicon leads (with screening length of few nanometres) till modern junctionless devices (Collinge, 2010). In far-from equilibrium conditions (i.e. high bias conditions), the standard screening lengths have to be complemented by an additional depletion length in the leads. The second solution (neglecting the leads) implies serious difficulties for the achievement of “overall charge neutrality”. In any case, a possible inaccuracy in the computation of the “overall charge neutrality” affects our ability to treat the time-dependent Coulomb correlation among electrons and, therefore, the requirement of “current conservation”. In conclusion, due to computational difficulties, modern electron transport simulators have to be implemented in small simulation boxes that imply important difficulties for providing accurate simulations of the DC or AC conductances of nanoscale devices.

In principle, the problem of excluding the leads from the simulation box, while retaining the lead-sample Coulomb correlation, can be solvable by providing adequate boundary conditions on each of the “opened” borders of the simulation box. However, such boundary conditions are not easily predictable. The standard boundary conditions found in the literature for nanoscale electron device simulators are based on specifying two conditions in each of the borders of the simulation box:

(Border-potential-BC).- The value of the scalar potential (or electric field) at the borders of the simulation box has to be specified. This condition is a direct consequence of the uniqueness theorem for the Poisson equation (Javid & Brown, 1963) which tells that such condition are enough to completely specify the solution of Poisson equation, when the charge inside the simulation box is perfectly determined (the reason for discarding the electromagnetic vector potential in nanoscale systems is discussed in Ref. Albareda, López, Cartoixà, Suñé & Oriols, 2010).

(Border-charge-BC).- Contrarily to what is needed for the uniqueness solution of the Poisson equation, the charge density inside the simulation box is uncertain because it depends on the electron injected from the borders of the simulation box. Therefore, any boundary condition algorithm has to include the information on the charge in the borders as an additional condition. In many cases, the electron injected on the borders depends, somehow, on the scalar potential there determined by the “Border-potential-BC” (and a fixed electrochemical potential). Therefore, a coupled system of boundary conditions appears.

Educated guesses for both boundary conditions (“Border-potential-BC” and “Border-charge-BC”) are present in the literature when describing nanoscale electron devices with simulation boxes large enough to include the leads. However, such boundary conditions are not applicable for small simulation boxes that exclude the leads. Here, we present a novel self-consistent and time-dependent definition of the boundary conditions for small simulation boxes (excluding most of the leads) that is able to capture the lead-sample Coulomb correlations (Albareda, López, Cartoixà, Suñé & Oriols, 2010).

### **3.2 Time-dependent boundary-conditions at the borders of the sample for overall charge neutrality**

As explained in the previous paragraphs, all boundary condition of electrons transport simulators are based on specifying the value of the scalar potential (or the electric field) in the

borders and the charge density there. Therefore, according to the labels of figure 7, we have to specify the values  $V_S(t)$  and  $V_D(t)$  for the “Border-potential-BC”, and  $\rho_S(t)$  and  $\rho_D(t)$  for the “Border-charge-BC”. Unfortunately, it is very difficult to provide an educated guess of the

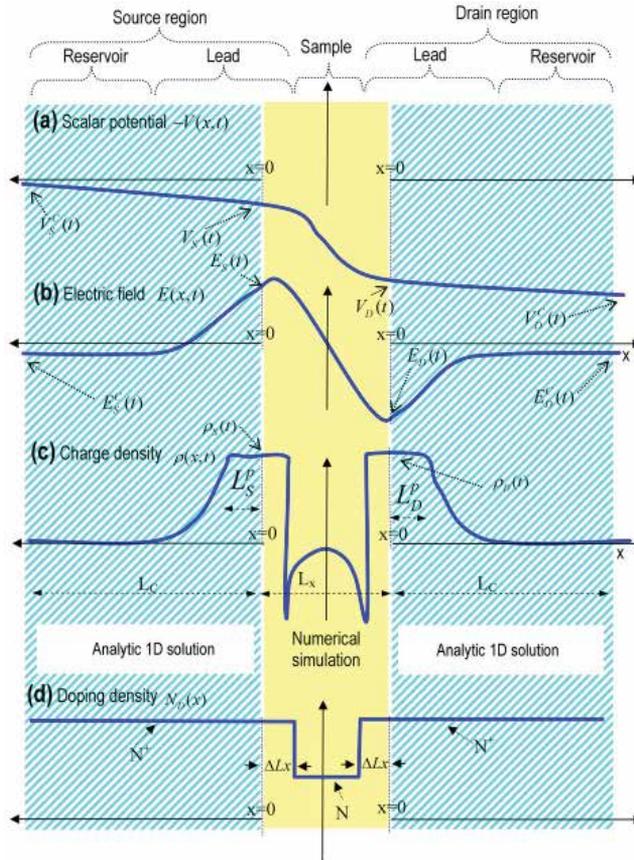


Fig. 7. Schematic description of the different parts of the electron device. In Ref. Albareda, López, Cartoixà, Suñé & Oriols, 2010, an analytical parametric 1D solution is deduced for the (blue) dashed region, while a numerical 3D solution is obtained in the (yellow) solid central region defined as the simulation box. Subsets refer to schematic representation of the (a) scalar potential, (b) electric field, (c) total charge density and (d) doping density. Reprinted with permission from G. Albareda et al., Phys. Rev. B. 79, 075315 (2009). ©Copyright 2009, American Physical Society.

scalar potential, the electric field or the charge density on the borders of a small simulation box that excludes the leads. For large simulation boxes, one can assume a known value of the electrochemical potential (deep inside the reservoir) that controls the electron injection. However, close to the active region, where there exists a far from equilibrium momentum distribution, the prediction of any value of the electrochemical potential is quiet inappropriate. From the results in Ref. Albareda, López, Cartoixà, Suñé & Oriols, 2010, we are able to translate the “Border-potential-BC” and “Border-charge-BC” discussed in section 3.1, for the borders of a small simulation box into simpler conditions deep inside the reservoirs. This is

the key point of our boundary conditions algorithm. In particular, the two new boundary conditions that we will impose at the limits of the simulation box,  $x = \mp L_C$  are (see Fig. 7):

“Deep-drift-BC”: We assume that the inelastic scattering mechanisms at, both, the source  $x \leq -L_C$  and the drain  $x \geq L_C$  reservoirs provides a non-equilibrium position-independent “thermal” distribution of electrons there (it is implicitly assumed that the contact length  $L_C$  is large enough and the temperature  $\Theta$  high enough so that inelastic scattering is relevant there). Such position-independent electron distribution is consistent with the “local” charge neutrality that implies a uniform electric field there. According to the Drude’s model, the electric fields there tend both to  $E_{S/D}^C(t) \rightarrow E_{S/D}^{drift}(t)$  [see expressions (11) and (12) of Ref. Albareda, López, Cartoixà, Suñé & Oriols, 2010].

“Deep-potential-BC”: We assume that electro-chemical potentials can be defined for the “thermal” distribution deep inside both reservoirs. As a consequence of the previous position-independent electron distribution deep inside the reservoirs, we can assume that the energy separation between such electro-chemical potential level and the bottom of the conduction band, in the drain and source reservoirs (at  $x = \mp L_C$ ) are equal. Therefore, the energy separation between the bottoms of the conduction bands at both reservoirs (which coincides with the separation of the electrochemical potentials) is equal to the difference of the external voltages. Thus,  $V_S^C(t) = 0$  and  $V_D^C(t) = V_{external}(t)$ .

These two conditions, “Deep-drift-BC” and “Deep-potential-BC” are quite reasonable deep inside the reservoirs. In fact, it can be shown that the numerical MC solution of the non-equilibrium BTE in a large simulation box provides exactly these results in the reservoirs (see Ref. Albareda, López, Cartoixà, Suñé & Oriols, 2010).

In order to translate the above “deep” boundary conditions into practical considerations on the simulation box borders, our algorithm couples the charge density, the electric field and the scalar potential to the injection model by taking into account the electrostatic interaction among the electrons within the active region and those in the leads (see Ref. Albareda, López, Cartoixà, Suñé & Oriols, 2010). Following this strategy, the amount of charge on the whole circuit can be set to zero, and thus “overall charge neutrality” and “current conservation” requirements are accomplished. The main approximation used to obtain the previous results is Drude’s law. Thus, our time-dependent boundary condition algorithm is only valid for frequencies below the inverse of the average electron scattering time (see Ref. Albareda, López, Cartoixà, Suñé & Oriols, 2010). In good reservoirs such frequencies are much higher than the THz range, which is high enough for most practical electronic applications.

The formulation of the previous boundary conditions, however, corresponds to a single-particle system. That is, we assume that all electrons are subjected to the same boundary conditions. Again, this is a simplification of the real many-particle problem. As we have shown in the previous section (see equations (10), (11) and (11)), every single electron “sees” its own electrostatic potential, electric field and charge distribution. Consequently, each electron should see its own boundary conditions. In the next section we extent these boundary conditions to a many-particle ones.

### 3.3 Extension of the boundary conditions to many-particle Hamiltonians

In the previous subsection we have presented a unique set of time-dependent boundary conditions for all electrons to account for overall charge neutrality and current conservation. Here we extend such results to a many-particle level where each electron has its own boundary conditions. Let us recall, that we are looking for solutions of the  $N(t)$  Poisson equations (11). Thus, we need to specify  $N(t)$  boundary conditions on the two opened border surfaces  $S^1$  and  $S^4$  (see figure 6) for the  $N(t)$  terms  $W_k(\vec{r}_1, \dots, \vec{r}_k, \dots, \vec{r}_{N(t)})$ .

In order to provide a clear notation for discussing the boundary conditions of  $W_k(\vec{r}_1, \dots, \vec{r}_k, \dots, \vec{r}_{N(t)})$ , we distinguish between the “source” vectors  $\{\vec{r}_1, \dots, \vec{r}_{k-1}, \vec{r}_{k+1}, \dots, \vec{r}_{N(t)}\}$  and the additional “observation” vector  $\vec{r}$  that runs over all space (Javid & Brown, 1963). In particular, the electrostatic potential that appears in the Hamiltonian (9) is defined as the value of the potential  $W_k(\vec{r}_1, \dots, \vec{r}_{k-1}, \vec{r}, \vec{r}_{k+1}, \dots, \vec{r}_{N(t)}, t)$  at the particular position  $\vec{r} = \vec{r}_k$ :

$$W_k(\vec{r}_1, \dots, \vec{r}_{k-1}, \vec{r}, \vec{r}_{k+1}, \dots, \vec{r}_{N(t)}, t) = W_k(\vec{r}_1, \dots, \vec{r}_{k-1}, \vec{r}, \vec{r}_{k+1}, \dots, \vec{r}_{N(t)}, t) \Big|_{\vec{r}=\vec{r}_k} \quad (18)$$

Our goal is to find an educated guess for all the  $N(t)$  terms  $W_k(\vec{r}_1, \dots, \vec{r}_{k-1}, \vec{r}, \vec{r}_{k+1}, \dots, \vec{r}_{N(t)}, t)$  at all “observation” points  $\vec{r} = \vec{r}_S$  and  $\vec{r} = \vec{r}_D$  on the surfaces  $S^1$  and  $S^4$ . The information of such boundary conditions comes from the value of the total voltage (due to internal and external electrons) at position  $\vec{r}_{S/D}$  and time  $t$ , that can be defined as the electrostatic potential associated to an additional probe charge  $q_{M+1}$  situated on that boundary,  $\vec{r}_{S/D} \equiv \vec{r}_{M+1} \in S^{4/1}$ . This potential can be then identified with the voltages  $V_{S/D}(t)$  defined in the previous subsection (see fig. 8), i.e.

$$V_{S/D}(t) \equiv \sum_{j=1}^M V(\vec{r}_{M+1}, \vec{r}_j) \Big|_{\vec{r}_{M+1}=\vec{r}_{S/D}} \quad (19)$$

where the expected restriction  $j \neq M+1$  is hidden in the limit of the sum. Once the relationship (19) is established, we can easily define the boundary conditions of any of the  $N(t)$  electrostatic potential  $W_k(\vec{r}_1, \dots, \vec{r}, \dots, \vec{r}_{N(t)})$  from the function  $V_{S/D}(t)$ . In particular, from (13), we know that:

$$W_k(\vec{r}_1, \dots, \vec{r}_{k-1}, \vec{r}, \vec{r}_{k+1}, \dots, \vec{r}_{N(t)}, t) \Big|_{\vec{r}=\vec{r}_{S/D}} = \sum_{\substack{j=1 \\ j \neq k}}^M V(\vec{r}_{S/D}, \vec{r}_j) = V_{S/D}(t) - V(\vec{r}_{S/D}, \vec{r}_k) \quad ; \quad l = 1, \dots, 6 \quad (20)$$

where  $V(\vec{r}_{S/D}, \vec{r}_k)$  is defined according to (3).

## 4. Monte Carlo solution of the many-particle open system Hamiltonian

Once we have introduced the many-particle open system Hamiltonian and its boundary conditions, we are ready to solve the electron transport problem. The classical description of the particle dynamics subjected to the many-particle Hamiltonian (10) can be computed by using the well-known Hamilton equations. In particular, we can obtain the (Newton like)

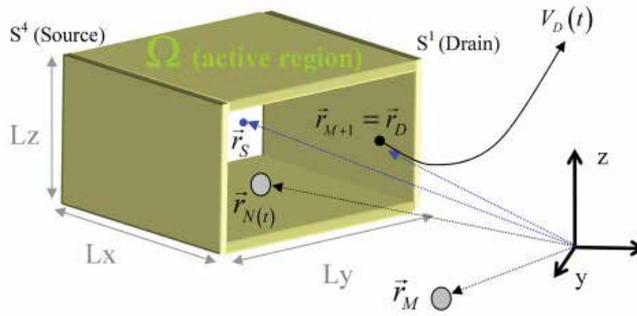


Fig. 8. The electrostatic potential  $V_D(t)$  (due to internal and external electrons) measured on the surface  $S^1$  at position  $\vec{r}_D$  and time  $t$  by an additional probe charge  $q_{M+1}$  situated on the boundary  $\vec{r}_D \equiv \vec{r}_{M+1} \in S^1$ .

description of the classical trajectory  $\vec{r}_i[t]$  in the real space through:

$$\frac{d\vec{p}_i[t]}{dt} = \left[ -\nabla_{\vec{r}_i} H(\vec{r}_1, \dots, \vec{r}_{N(t)}, \vec{p}_1, \dots, \vec{p}_{N(t)}, t) \right]_{\vec{r}_1=\vec{r}_1[t], \dots, \vec{p}_{N(t)}=\vec{p}_{N(t)}[t]} \quad (21a)$$

$$\frac{d\vec{r}_i[t]}{dt} = \left[ \nabla_{\vec{p}_i} H(\vec{r}_1, \dots, \vec{r}_{N(t)}, \vec{p}_1, \dots, \vec{p}_{N(t)}, t) \right]_{\vec{r}_1=\vec{r}_1[t], \dots, \vec{p}_{N(t)}=\vec{p}_{N(t)}[t]} \quad (21b)$$

For the many-particle Hamiltonian studied here, expression (21b) gives the trivial result  $m \cdot \vec{v}_i[t] = \vec{p}_i[t]$ , while the evaluation of expression (21a) requires a more detailed development. We know that the  $\vec{r}_i$ -gradient of the exact many-particle Hamiltonian (10) can be written as:

$$\left[ \nabla_{\vec{r}_i} H \right]_{\vec{R}=\vec{R}[t]} = \left[ \nabla_{\vec{r}_i} \sum_{k=1}^{N(t)} \left\{ e \cdot W_k(\vec{r}_1, \dots, \vec{r}_{N(t)}, t) - \frac{1}{2} \sum_{\substack{j=1 \\ j \neq k}}^{N(t)} e \cdot V(\vec{r}_k, \vec{r}_j) \right\} \right]_{\vec{R}=\vec{R}[t]} \quad (22)$$

We define the multi-dimensional vector  $\vec{R} = (\vec{r}_1, \dots, \vec{r}_{N(t)})$  to account, in a compact way, for the classical trajectories of  $N(t)$  electrons  $\vec{R}[t] = (\vec{r}_1[t], \dots, \vec{r}_{N(t)}[t])$ . Substituting the definition of  $W_k(\vec{r}_1, \dots, \vec{r}_{N(t)}, t)$  of expression (13) into equation (22), we find:

$$\left[ \nabla_{\vec{r}_i} H \right]_{\vec{R}=\vec{R}[t]} = \left[ \nabla_{\vec{r}_i} \left\{ 2 \sum_{\substack{j=1 \\ j \neq i}}^{N(t)} e V(\vec{r}_j, \vec{r}_i) + \sum_{j=N(t)+1}^M e V(\vec{r}_j, \vec{r}_i) \right\} - \nabla_{\vec{r}_i} \sum_{\substack{j=1 \\ j \neq i}}^{N(t)} e \cdot V(\vec{r}_j, \vec{r}_i) \right]_{\vec{R}=\vec{R}[t]} \quad (23)$$

Now, from expressions (13) and (23), we realize that:

$$\left[ \nabla_{\vec{r}_i} H \right]_{\vec{R}=\vec{R}[t]} = \left[ \nabla_{\vec{r}_i} W_i(\vec{r}_1, \dots, \vec{r}_{N(t)}) \right]_{\vec{R}=\vec{R}[t]} \quad (24)$$

Only the term  $W_i(\vec{r}_1, \dots, \vec{r}_{N(t)})$  of the whole Hamiltonian (10) becomes relevant for a classical description of the  $i$ -particle. In fact, since we only evaluate a  $\vec{r}_i$ -gradient, the rest of particle positions can be evaluated at their particular value at time  $t$ , i.e.  $\vec{r}_k \rightarrow \vec{r}_k[t]$  for all  $k \neq i$ .

Therefore, we define the single-particle potential  $\bar{W}_i(\vec{r}_i, t)$  from the many-particle potential as:

$$\bar{W}_i(\vec{r}_i, t) = W_i(\vec{r}_1[t], \dots, \vec{r}_{i-1}[t], \vec{r}_i, \vec{r}_{i+1}[t], \dots, \vec{r}_{N(t)}[t]). \quad (25)$$

We use a “hat” to differentiate the (time-dependent) single-particle electrostatic potential from the many-particle potential. Each i-term of the single-particle electrostatic potential,  $\bar{W}_i(\vec{r}_i, t)$ , is a solution of one particular 3D-Poisson equation:

$$\nabla_{\vec{r}_i}^2 (\epsilon(\vec{r}_i) \cdot \bar{W}_i(\vec{r}_i, t)) = \bar{\rho}_i(\vec{r}_i, t), \quad (26)$$

where the single-particle charge density is defined as:

$$\bar{\rho}_i(\vec{r}_i, t) = \sum_{\substack{j=1 \\ j \neq i}}^{N(t)} e \delta(\vec{r}_i - \vec{r}_j[t]), \quad (27)$$

and the boundary conditions are adapted here as:

$$\bar{W}_i(\vec{r}_i, t) \Big|_{\vec{r}_i = \vec{r}_{S/D}} = V_{S/D}(t) - V(\vec{r}_{S/D}, \vec{r}_i[t]). \quad (28)$$

Let us remind that expressions (25), (26) and (27) together with the boundary conditions in (28), provide an exact treatment of the many-particle correlations in classical scenarios. The  $N(t)$  Newton equations are coupled by  $N(t)$  Poisson equations. Therefore, the many-particle Hamiltonian of (10) can be written exactly (without mean-field approximation) as:

$$H(\vec{r}_1, \dots, \vec{r}_{N(t)}, \vec{p}_1, \dots, \vec{p}_{N(t)}, t) = \sum_{k=1}^{N(t)} \{K(\vec{p}_k) + e \cdot \bar{W}_k(\vec{r}_k, t)\}. \quad (29)$$

It is important to recall here that, although electron dynamics within our open system is *deterministically* described by the many-particle Hamiltonian (10) supplied with the Hamilton equations (21), our simulations will be subject to an *stochastic injection* of electrons describing how (i.e. in which position, time and momentum) electrons enter the simulated region. Indeed, our approach to electron transport ultimately constitutes an statistical problem. Due to computational limitations, we have been forced to reduce the degrees of freedom of our system. Since we can only describe a very reduced number of variables in a very reduced region of space (an open system representing the active region of an electron device), we are obliged to deal with an essentially uncertain environment. The injection process, is then the responsible of coupling an statistical *external* environment to our “deterministic” simulation box, and thus, it is also the main responsible of converting the information that we have on the dynamics occurring inside the simulation region into something statistical (Albareda, López, Cartoixà, Suñé & Oriols, 2010). It is in this regard that we can classify our approach to electron transport as a MC technique. Nonetheless, as we have already announced, our many-electron method applied to semiclassical devices cannot be considered a solution of the BTE because the latter is developed at a classical mean-field level. The term  $\bar{W}_k(\vec{r}_k, t)$  in the Hamiltonian of expression (29) means that each particle “sees” its own electrostatic potential which is different to that of the others. Apart from the scattering rates, this is the fundamental difference between our “many-electron” method applied to classical transport and the standard MC solution of the BTE method for electron devices.

Let us just mention here that the previous algorithm developed to describe classical electron transport at a many-particle level can be easily generalized to quantum systems by means of an original formalism based on quantum trajectories (Oriols, 2007; Oriols & Mompert, 2011). In particular, recent efforts have made it possible to demonstrate the viability to develop a quantum trajectory-based approach to electron transport with many-particle correlations (Albareda, 2010; Albareda, López, Cartoixà, Suñé & Oriols, 2010; Albareda, Suñé & Oriols, 2009). We have named this simulator BITLLES.

## 5. Numerical example: Many-particle transport in the channel of quantum wire DG-FETs with charged atomistic impurities

Up to now we have been focused mainly on the theoretical aspects of our MC approach to electron transport. Sections 2, 3 and 4 constitute the keystone pieces for the development of a versatile simulation tool capable of describing semi-classical electron transport including Coulomb correlations at a many-particle level. The aim of this section is to present an example of the capabilities of such a simulator to predict certain relevant aspects of future nanoscale electron devices. In particular, we want to highlight the importance of accurately accounting for (time-dependent) Coulomb correlations among (transport) electrons in the analysis of discrete doping induced fluctuations.

Differences in number and position of dopant atoms in sub-10nm channel devices will produce important variations on the devices' microscopic behavior, and consequently, the variability of macroscopic parameters such as drive current or threshold voltage will increase. This particular phenomenon is known as discrete dopant induced fluctuations, and constitutes one of the most reported causes of variations from sample to sample in electron devices characteristics (coming from the atomistic nature of mater). We study here the effect of single ionized dopants on the performance of a quantum wire double-gate FET (QWDG-FET), mainly when its lateral dimensions approach the effective cross section of the charged impurities. We find that neglecting the (time-dependent) Coulomb correlations among (transport) electrons can lead to misleading predictions of devices behavior (Albareda, Saura, Oriols & Suñé, 2010).

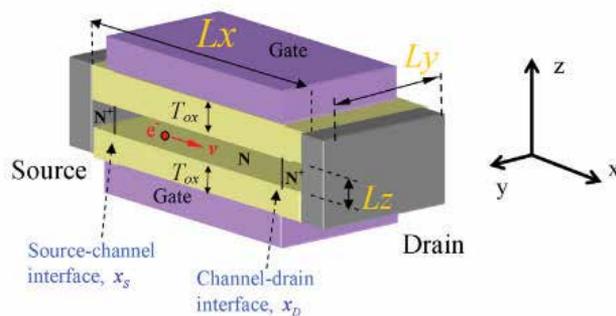


Fig. 9. Schematic representation of the quantum wire double gate FET.

### 5.1 Device Characteristics and Simulation Details

The structure of the simulated QWDG-FET is described in figure 9. Two highly N doped Si contacts ( $N^+ = 2 \cdot 10^{19} \text{ cm}^{-3}$ ) are connected to an intrinsic Si channel with lateral dimensions

$L_y = 5nm$  and  $L_z = 2nm$ . Such dimensions originate quantum confinement in the lateral directions (a quantum wire), not only reducing the degrees of freedom of the system but also inducing volume inversion within the channel (Balestra et al., 1987). In this regard, the electrostatic blockade generated by the ionized dopants is expected to be favored when impurities are distributed mainly in the center of the channel cross section. At the same time, the length of the quantum wire is  $10nm$ . This geometry results in a volume of only  $100nm^3$ , so that the number of interacting electrons in the channel is of the order of 10. Under such special conditions, the importance of the correlation among electrons is expected to be particularly relevant.

Electron transport in the “x” direction (from source to drain) takes place along a Silicon (100) oriented channel, at room temperature. In particular, the electron mass is taken according to the six equivalent ellipsoidal constant energy valleys of the silicon band structure (Jacoboni & Reggiani, 1983; Oriols et al., 2007). The effective masses of the ellipsoids are  $m_l^* = 0.9163 m_0$  and  $m_t^* = 0.1905 m_0$  with  $m_0$  being the free electron mass (Jacoboni & Reggiani, 1983). As commented above, the lateral dimensions of the Si channel  $L_z$  and  $L_y$  are both small enough, so that the active region behaves as a 1D system and the energy of an electron in one particular valley is  $E = \hbar^2 k_x^2 / (2m_l) + E_{1D}^q$ , where  $E_{1D}^q = \hbar^2 \pi^2 / (2m_t L_y^2) + \hbar^2 \pi^2 / (2m_l L_z^2)$  represents the minimum energy of the first sub-band, whose value is  $E_{1D}^q = 0.182eV$  for  $L_z = 2nm$  and  $L_y = 5nm$ . The energies of the next lowest sub-bands ( $E_{1D}^q = 0.418eV$  or  $E_{1D}^q = 0.489eV$ ) are assumed to be high enough to keep a single band simulation sufficiently accurate. Therefore, we use a 3D Poisson solver to deal with the device electrostatics, but a 1D algorithm to describe the velocity of each electron in the “x” direction. Due to the lateral electron confinement, the velocities in “y” and “z” directions are zero<sup>6</sup>. This is an exact result for describing electron confinement in the rectangular structure of figure 9 when the e-e and e-i are not taken into account. The explicit consideration of the effect of e-e and e-i correlations on the electron confinement (energy levels) is an extremely complicated issue within the many-particle strategy developed here, which considers one different scalar potential for each electron. The reader can find more information on the simulation details in Ref. Albareda, Saura, Oriols & Suñé, 2010.

## 5.2 Analysis of discrete doping induced fluctuations

In order to highlight the importance of taking into account the time-dependent e-e and e-i correlations, we will compare some results with those obtained through a single-particle mean-field approach discussed in Ref. Albareda, Saura, Oriols & Suñé, 2010. In this regard, we will refer to many-particle results to describe the simulation performed with the algorithms that require solving  $N(t)$  Poisson equations with  $N(t)$  charge densities [expressions (26),(27),(28)] at each time step. Alternatively, we will refer to the time-independent single-particle approximation to the more simplistic (though usual) approach that consists in solving a single time-independent Poisson equation [expressions (A1) and (A2) in Ref. Albareda, Saura, Oriols & Suñé, 2010 for all electrons at each time step of the simulation.

<sup>6</sup> We assume that the electron velocity is equal to zero in the lateral directions where there is energy confinement. This is a reasonable assumption that can be formally justified (see Ref. Oriols, 2007) when the probability presence in that direction does not change with time. The main approximation here is assuming that the time dependence of the wave function involves only one quantized energy in the mentioned direction. We define the geometry of the QWDG-FET to support these approximations.

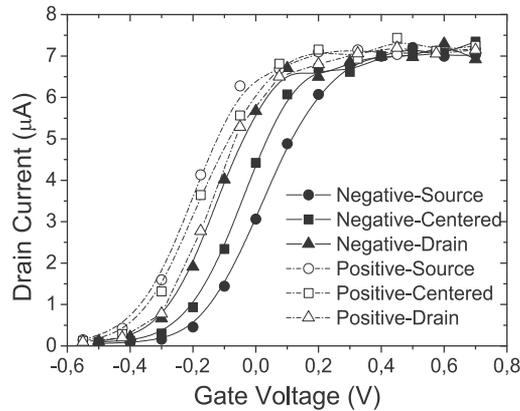


Fig. 10. Average drain current at  $V_{Drain} = 1V$  as a function of the gate voltage for positive/negative impurities located at different places along the channel. Reprinted with permission from G. Albareda et al., J. Appl. Phys. 108, 043706 (2010). ©Copyright 2010, American Institute of Physics.

Let us start the discussion with the study of threshold voltage ( $V_T$ ) fluctuations, a well known effect related to discrete doping induced fluctuations in MOSFETs (Asenov, 1999; Asenov et al., 2003; Gross et al., 2002; Millar et al., 2008; Reid et al., 2009; Vasileska & Ahmed, 2005). We first analyze this phenomenon in the QWDG-FET for both positive and negative impurities. Fig. 10 shows the value of the mean current as a function of the applied gate voltage (transfer characteristic) in the saturation region ( $V_{Drain} = 1V$ ). While negative ions induce a shift of the threshold voltage towards higher values, positively charged impurities shift it down to lower values. The explanation of such a behavior is quite simple. Since the majority carriers are electrons, positive charged impurities introduce a potential well that favors the flow of the current, while negative impurities appear as potential barriers which tend to block the transmission of electrons. A dependence of the saturation threshold voltage on the position of the impurities along the channel can also be observed. As a negative (positive) dopant is displaced from drain to source, the threshold voltage is increased (decreased) in a non-linear way due to an increment of the height (depth) of the induced potential deformation that is less and less masked by the applied drain voltage (Albareda, Saura, Oriols & Suñé, 2010).

Next, we show how negative dopants placed in the channel of a QWDG-FET induce significant changes in the spatial distribution of the current-density across the channel section. We consider the steady state current corresponding to a fixed bias point ( $V_{Gate} = 0V$ ;  $V_{Drain} = 0.5V$ ) and analyze the spatial distribution of the current in the channel cross section. Since we deal with a confined electron system under stationary conditions, the continuity equation reduces to  $\vec{\nabla} \cdot J_x = 0$  and consequently the spatial distribution of the current-density,  $J_x$ , is the same in any section along the transistor channel. In figure 11 we present the current-density distribution when a negative impurity is located at the source-channel interface. As it can be observed, the potential barrier produced by the dopant induces an important deformation of the spatial current density distribution, pushing carriers away from its location. This is a common result to both single- and a many-particle treatment of electron transport. However, differences on the magnitude of the current-density between Fig. 11.a) and Fig.

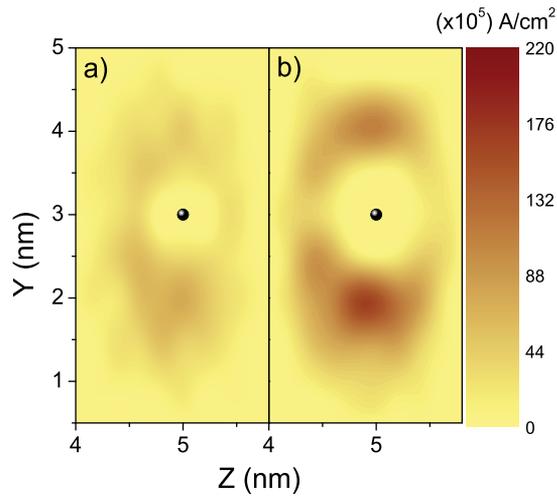


Fig. 11. Current density across the channel section when the negative impurity is placed at the center of the channel length. In a) the results correspond to a many-particle treatment of the system. In b) the results have been computed within a single-particle mean-field approach discussed in Ref. (Albareda, Saura, Oriols & Suñé, 2010). Reprinted with permission from G. Albareda et al., *J. Appl. Phys.* 108, 043706 (2010). ©Copyright 2010, American Institute of Physics.

11.b) are only attributable to fundamental differences among both treatments. Although from a single-particle point of view, a one-by-one electron energy conservation must be accomplished, many-particle transport implies a much looser restriction on the energy of the carriers.

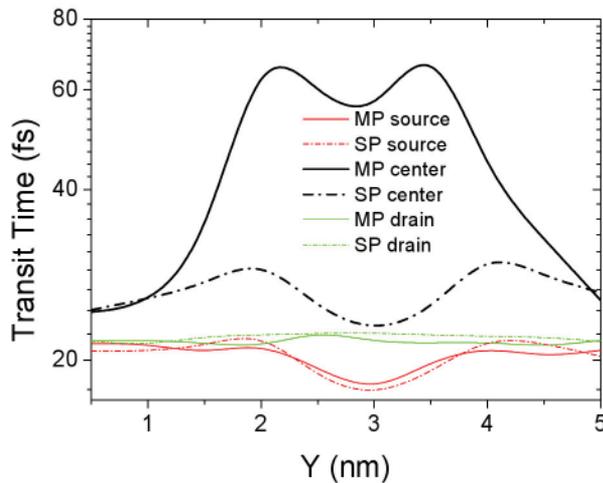


Fig. 12. Spatial distribution of the transit times along the  $y$  direction (centred in  $z$ ) when a negatively charged impurity is placed at different places of the channel. Reprinted with permission from G. Albareda et al., *J. Appl. Phys.* 108, 043706 (2010). ©Copyright 2010, American Institute of Physics.

In figure 12 we represent the distribution of the transit times along the  $y$  (centered in  $z$ ) direction. If all the traversing carriers had the same total energy, one would expect to find the largest transit times concentrated around the impurity location, where the potential barrier is higher (see Fig. 12). Nevertheless, since the injected carriers are energetically spread according to Fermi statistics, only the fastest electrons (the most energetic ones) are able to achieve the drain contact across the top of the barrier. Therefore, a minimum of the transit time is found at the location of the impurity atom. When the dopant is placed at the source-channel interface, the largest transit times appear away from the impurity and the minimum above the dopant becomes absolute. Although both, single- and many-particle simulations give similar overall results in this case, some discrepancies can be appreciated due to an energy exchange among the different regions of the channel. When the impurity is placed in the center of the channel, the transit times increase drastically up to 60 fs. Although the shape of the scalar potential is the responsible of the important increment in both the single-particle and the many-particle transit times results, electron-electron interactions play a crucial role in explaining the important differences between these two treatments. Since the spatial integral of the many-particle averaged transit times along the  $y$  and  $z$  directions diverges significantly from its single-particle counterpart, it can be inferred that the exchange of energy is produced not only among the electrons crossing the channel but also between them and those being backscattered (Albareda, Saura, Oriols & Suñé, 2010). Indeed, if the energy transfer would only involve electrons crossing the device, their total energy would remain unchanged, and thus their averaged transit time would be identical to that found for the single-particle approach. On the contrary, the mixture of energy exchange among the traversing electrons and among traversing and backscattered electrons give rise to a non conserving averaged transit time.

## 6. Conclusions

In this chapter we have presented a semi-classical MC approach to electron transport at the nanoscale without assuming any mean-field or perturbative approximation to describe the Coulomb interaction among transport electrons. Sections 2, 3 and 4 constitute the theoretical core of our semi-classical MC approach. In section 2, a many-particle Hamiltonian for  $N(t)$  electrons inside an open system has been developed. Departing from the exact Hamiltonian of a whole closed circuit, using a single-band effective mass approximation we are capable of developing a many-particle Hamiltonian (10) constituted by a sum of  $N(t)$  electrostatic potentials,  $W_k(\vec{r}_1, \dots, \vec{r}_k, \dots, \vec{r}_{N(t)})$ , solution of  $N(t)$  different Poisson equations (11). In section 3, we have presented a novel boundary conditions algorithm capable of describing the Coulomb correlations among electrons inside and outside the open system without significantly increasing the simulated degrees of freedom. In terms of analytical expressions describing the charge density, the electric field and the scalar potential along the leads and reservoirs, we can transfer the assumptions about the boundary conditions at the borders of a small simulation boxes into the simpler specifications of the boundary conditions deep inside the reservoirs. Our boundary conditions algorithm is able to discuss far from equilibrium situations where depletion lengths in the leads have to be added to standard screening. More over, the frequency-dependent correlations included into our boundary conditions algorithm, due to sample-lead Coulomb interaction, allow us to investigate the computation of (zero-frequency and high-frequency) current fluctuations beyond the standard external zero impedance assumption (i.e. most of the computations of current fluctuations in

electron devices assume that the voltage applied in the simulation box is a non-fluctuating quantity). In section 4, we have presented a classical solution of the many-particle open system Hamiltonian supplied with our many-particle boundary conditions. The solution is obtained via a coupled system of Newton-like equations with a different electric field for each particle, and constitutes a many-particle generalization of the MC solution of the semi-classical single-particle Boltzmann distribution. In the last section, our many-particle approach to electron transport has been applied to predict the behavior of some characteristics of a QWDG-FET under the presence of discrete impurities. We have revealed the significant impact of the sign and position of the impurity along the transistor channel on the threshold voltage, but more importantly, a comparison with more standard simulations which assume a time-independent mean-field approximation has allowed us to highlight the importance of an accurate treatment of the e-e interactions in the study of discrete doping induced fluctuations in nanometer scale devices. Finally, let us emphasize that many efforts are being devoted in the literature to improve the treatment of electron (and electron-atom) correlations on the description of the band structure of nanoscale devices (i.e. the ground-state at equilibrium conditions). Contrarily, in this work we open a new path to study the effects of electron (and electron-impurity) correlations in the current measured in nanoscale electron devices under (applied bias) far from equilibrium conditions.

Finally, let us mention that the presented technique can be also extended to describe quantum transport by means of bohmian trajectories (Alarcon & Oriols, 2009; Albareda, 2010; Albareda, López, Cartoixà, Suñé & Oriols, 2010; Albareda, Suñé & Oriols, 2009; Oriols, 2007). In this regard, a versatile (classical and quantum) many-particle approach to electron transport, called BITLLES, capable of reproducing DC, AC and noise performance in quantum scenarios has been already developed (Albareda, 2010; Oriols & Mompert, 2011).

## Acknowledgments

We would like to thank doctors D. Pardo, T. González and J. Mateos from Universidad de Salamanca. The MC method presented in this work is an evolution of their technique towards simulations tools with a full treatment of Coulomb correlations. This work has been partially supported by the Ministerio de Ciencia e Innovación under Project No. TEC2009-06986 and by the DURSI of the Generalitat de Catalunya under Contract No. 2009SGR783.

## 7. References

- Alarcon, A. & Oriols, X. (2009). Computation of quantum electron transport with local current conservation using quantum trajectories, *Journal of Statistical Mechanics: Theory and Experiment* 2009: P01051.
- Albareda, G. (2010). *Classical and Quantum Trajectory-based Approaches to Electron Transport with full Coulomb Correlations*, PhD thesis, Universitat Autònoma de Barcelona.
- Albareda, G., Jiménez, D. & Oriols, X. (2009). Intrinsic noise in aggressively scaled field-effect transistors, *Journal of Statistical Mechanics* 2009: P01044.
- Albareda, G., López, H., Cartoixà, X., Suñé, J. & Oriols, X. (2010). Time-dependent boundary conditions with lead-sample coulomb correlations: Application to classical and quantum nanoscale electron device simulators, *Physical Review B* 82(8): 085301.

- Albareda, G., Saura, X., Oriols, X. & Suñé, J. (2010). Many-particle transport in the channel of quantum wire double-gate field-effect transistors with charged atomistic impurities, *Journal of Applied Physics* 108(4): 043706.
- Albareda, G., Suñé, J. & Oriols, X. (2009). Many-particle hamiltonian for open systems with full coulomb interaction: Application to classical and quantum time-dependent simulations of nanoscale electron devices, *Physical Review B* 79(7): 075315.
- Alexander, C., Brown, A., Watling, J. & Asenov, A. (2005). Impact of single charge trapping in nano-mosfets-electrostatics versus transport effects, *Nanotechnology, IEEE Transactions on* 4(3): 339.
- Alexander, C., Roy, G. & Asenov, A. (2008). Random-dopant-induced drain current variation in nano-MOSFETs: A three-dimensional self-consistent monte carlo simulation study using "ab initio" ionized impurity scattering, *IEEE Transactions on Electron Devices* 55(11): 3251.
- Asenov, A. (1999). Random dopant induced threshold voltage lowering and fluctuations in sub 50 nm MOSFETs: a statistical 3d 'atomistic' simulation study, *Nanotechnology* 10(2): 153.
- Asenov, A., Brown, A. R., Davies, J. H., Kaya, S. & Slavcheva, G. (2003). Simulation of intrinsic parameter fluctuations in decananometer and nanometer-scale MOSFETs, *IEEE Transactions on Electron Devices* 50(9): 1837.
- Asenov, A., Brown, A., Roy, G., Cheng, B., Alexander, C., Riddet, C., Kovac, U., Martinez, A., Seoane, N. & Roy, S. (2009). Simulation of statistical variability in nano-cmos transistors using drift-diffusion, monte carlo and non-equilibrium green's function techniques, *Journal of Computational Electronics* 8: 349.
- Babiker, S., Asenov, A., Cameron, N. & Beaumont, S. P. (1996). Simple approach to include external resistances in the monte carlo simulation of mesfets and hemts, *Transactions on Electron Devices, IEEE* 43: 2032.
- Balestra, F., Cristoloveanu, S., Benachir, M., Brini, J. & Elewa, T. (1987). Double-gate silicon-on-insulator transistor with volume inversion: A new device with greatly enhanced performance, *IEEE Electron Device Letters* 8: 410.
- Barraud, S., Dollfus, P., Galdin, S. & Hesto, P. (2002). Short-range and long-range coulomb interactions for 3d monte carlo device simulation with discrete impurity distribution, *Solid-State Electronics* 46: 1061.
- Blanter, Y. M. & Büttiker, M. (2000). Shot noise in mesoscopic conductors, *Physics Reports* 336(1-2): 1.
- Boltzmann, L. W. (1872). Weitere studien uber das wärmegewicht unter gasmolekülen, *Ber. Wien. Akad.* 66: 275.
- Brandbyge, M., Mozos, J., Ordejón, P., Taylor, J. & Stokbro, K. (2002). Density-functional method for nonequilibrium electron transport, *Physical Review B* 65(16): 165401.
- Bulashenko, O. M., Mateos, J., Pardo, D., González, T., Reggiani, L. & Rubí, J. M. (1998). Electron-number statistics and shot-noise suppression by coulomb correlation in nondegenerate ballistic transport, *Physical Review B* 57(3): 1366.
- Collinge, J. P. e. a. (2010). Nanowire transistors without junctions, *Nature Nanotechnology* 5: 225.
- Conwell, E. M. (1967). *High field transport in semiconductors, Solid-state Phys. Suppl.* 9, New York Academic.
- Datta, S. (1995). *Electronic transport in mesoscopic systems*, Cambridge University Press.
- Datta, S. (2005). *Quantum Transport: Atom to Transistor*, Cambridge University Press.

- De Mari, A. (1968). An accurate numerical steady-state one-dimensional solution of the p-n junction, *Solid-State Electronics* 11: 33.
- Di Ventra, M. (2008). *Electrical Transport in Nanoscale Systems*, first edn, Cambridge University Press, The Edinburgh Building, Cambridge CB2 8RU, UK.
- Fischetti, M. V. & Laux, S. E. (2001). Long-range coulomb interactions in small si devices. part i: Performance and reliability, *Journal of Applied Physics* 89: 1205.
- Gomila, G., Cantalapiedra, I. R., González, T. & Reggiani, L. (2002). Semiclassical theory of shot noise in ballistic  $n^+ - i - n^+$  semiconductor structures: Relevance of pauli and long-range coulomb correlations, *Physical Review B* 66(7): 075302.
- Gonzalez, T., Bulashenko, O. M., Mateos, J., Pardo, D. & Reggiani, L. (1997). Effect of long-range coulomb interaction on shot-noise suppression in ballistic transport, *Physical Review B* 56: 6424.
- Gonzalez, T. & Pardo, D. (1993). Ensemble monte carlo with poisson solver for the study of current fluctuations in homogeneous gaas structures, *Journal of Applied Physics* 73: 7453.
- Gonzalez, T. & Pardo, D. (1996). Physical models of ohmic contact for monte carlo device simulation, *Solid-State Electronics* 39: 555.
- Gross, W. J., Vasileska, D. & Ferry, D. K. (1999). A novel approach for introducing the electron-electron and electron-impurity interactions in particle-based simulations, *Electron Device Letters, IEEE* 20(9): 463.
- Gross, W. J., Vasileska, D. & Ferry, D. K. (2000a). 3d simulations of ultra-small MOSFETs with real-space treatment of the electron-electron and electron-ion interactions, *VLSI Design* 10: 437.
- Gross, W. J., Vasileska, D. & Ferry, D. K. (2000b). Ultrasmall MOSFETs: the importance of the full coulomb interaction on device characteristics, *Transactions on Electron Devices, IEEE* 47(10): 1831.
- Gross, W. J., Vasileska, D. & Ferry, D. K. (2002). Three-dimensional simulations of ultrasmall metal-oxide-semiconductor field-effect transistors: The role of the discrete impurities on the device terminal characteristics, *Journal of Applied Physics* 91: 3737.
- Gummel, H. (1964). A self-consistent iterative scheme for one-dimensional steady state transistor calculations, *Transaction on Electron Devices, IEEE* 11: 455.
- Jacoboni, C. & Lugli, P. (1989). *The Monte Carlo Method for Semiconductor Device Simulation*, Springer-Verlag Wien.
- Jacoboni, C. & Reggiani, L. (1983). The monte carlo method for the solution of charge transport in semiconductors with applications to covalent materials, *Review of Modern Physics* 55(3): 645.
- Javid, M. & Brown, P. M. (1963). *Field Analysis and Electromagnetics*, McGraw-Hill.
- Kurosawa, T. (1966). Proceeding of the international conference on the physics of semiconductors, *Journal of the Physical Society of Japan Supplement* 21: 424.
- Landauer, R. (1992). Conductance from transmission: common sense points, *Physica Scripta* T42: 110.
- Lundstrom, M. & Guo, J. (2006). *Nanoscale Transistors: Device Physics, Modeling and Simulation*, Springer Science.
- Millar, C., Reid, D., Roy, G., Roy, S. & Asenov, A. (2008). Accurate statistical description of random dopant-induced threshold voltage variability, *Electron Device Letters, IEEE* 29: 946.

- Oriols, X. (2007). Quantum-trajectory approach to time-dependent transport in mesoscopic systems with electron-electron interactions, *Physical Review Letters* 98(6): 066803.
- Oriols, X., Fernández-Díaz, E., Álvarez, A. & Alarcón, A. (2007). An electron injection model for time-dependent simulators of nanoscale devices with electron confinement: Application to the comparison of the intrinsic noise of 3d-, 2d- and 1d-ballistic transistors, *Solid-State Electronics* 51: 306.
- Oriols, X. & Mompert, J. (Unpublished). *Applied Bohmian Mechanics: From Nanoscale Systems to Cosmology*, Pan Stanford.
- Ramey, S. M. & Ferry, D. K. (2003). A new model for including discrete dopant ions into monte carlo simulations, *Transactions on Nanotechnology, IEEE* 2: 193.
- Reid, D., Millar, C., Roy, G., Roy, S. & Asenov, A. (2009). Analysis of threshold voltage distribution due to random dopants: A 100.000 sample 3d simulation study, *Transactions on Electron Devices, IEEE* 56: 2255.
- Reitz, J. R., Milford, F. J. & Christy, R. W. (1992). *Foundations of electromagnetic theory*, Addison-Wesley.
- Reklaitis, A. & Reggiani, L. (1999). Monte carlo study of shot-noise suppression in semiconductor heterostructure diodes, *Physical Review B* 60(16): 11683.
- Riddet, C., Brown, A. R., Roy, S. & Asenov, A. (2008). Boundary conditions for density gradient corrections in 3d monte carlo simulations, *Journal of Computational Electronics* 7: 231.
- Scharfetter, D. & Gummel, H. (1969). Large-signal analysis of a silicon read diode oscillator, *Transaction on Electron Devices, IEEE* 16: 64.
- Sverdlov, V., Ungersboek, E., Kpsina, H. & Selberherr, S. (2008). Current transport models for nanoscale semiconductor devices, *Materials Science and Engineering: Reports* 58: 228.
- Taylor, J., Guo, H. & Wang, J. (2001). Ab initio modeling of quantum transport properties of molecular electronic devices, *Physical Review B* 63: 245407.
- Thijssen, J. M. (2003). *Computational Physics*, Cambridge University Press.
- Vasileska, D. & Ahmed, S. (2005). Narrow-width soi devices: the role of quantum-mechanical size quantization effect and unintentional doping on the device operation, *Electron Devices, IEEE Transactions on* 52(2): 227 – 236.
- Wordelman, C. J. & Ravaioli, U. (2000). Integration of a particle-particle-particle-mesh algorithm with the ensemble monte carlo method for the simulation of ultra-small semiconductor devices, *Transaction on Electron Devices, IEEE* 47: 410.

# Monte-Carlo Simulation in Electron Microscopy and Spectroscopy

Vladimír Starý

*Czech Technical University in Prague,  
Czech Republic*

## 1. Introduction

In electron microscopy, spectroscopy and microanalysis, knowledge of certain quantities is very often needed for proper analytical measurement. Unfortunately, the real values of some of these quantities can only be roughly estimated for two reasons: the complicated process of electron and phonon transport through the matter can be only very approximately described by the analytical theory, and the experimental measurement of some quantities for proper evaluation of experiments is hardly possible. In this case, a Monte-Carlo simulation (MC) can give us reasonable results (see reviews (Berger, 1963), (Binder, 1979), (Joy, 1995), (Dapor, 2003)). As the main motivations of Monte-Carlo simulations in electron microscopy and spectroscopy we can put

- a. calculation of inelastic mean free path (IMFP) of electrons in matter,
- b. prognostics of the physical processes results.

The reliability of Monte-Carlo models is usually checked by comparison with experimental results; unfortunately, some quantities can be measured only indirectly or with some experimental problems. The basic problems lie

- a. in selecting suitable quantities, which can be at best directly measured and the results of calculations can be compared with them;
- b. in maximizing the range of these quantities for which the checking is valid (electron energy, atomic number, thickness of surface film, etc.)

When this method was first used, due to the low speed of computers, the multiple scattering type of calculation was usually used, and the relatively long parts of the path were simulated simultaneously using averaging of scattering effects. Nowadays, so-called single scattering models are employed, where each scattering event is calculated individually. In Monte-Carlo code, both the formulas and tables of values necessary for calculation can be used. Because of necessity of interpolation of values between values given by tables, the formulas are preferred.

In the energy range used for electron microscopy, spectroscopy and microanalysis (i.e., usually 0.1 to 300 keV), various models are used in MC codes for describing the basic interactions of electrons with atoms - elastic and inelastic collisions and other interactions of electrons with materials.

## 2. Description of the physical interactions of electrons with material

Interaction of particle  $A_1$  with nucleus  $A_2$  is described generally by formula

$$A_1 + A_2 \Rightarrow A_3 + A_4 + Q, \quad (1)$$

where  $A_3$  is the outgoing particle,  $A_4$  is resulting nucleus and  $Q$  is emitted energy.

### 2.1 Elastic scattering

For elastic scattering, the simulation started using the Rutherford formula (usually with unscreened or screened nucleus charge and sometimes with relativistic correction). Now more exact calculation of differential cross sections is provided by using the static field approximation of atomic potential (Dirac-Hartree-Fock-Slater, Thomas-Fermi-Dirac, etc.) with relativistic partial wave analysis (e.g., (Salvat & Mayol, 1993), (Mayol & Salvat, 1997), (Salvat et al., 2005)). Moreover, the Hartree-Fock-Wigner-Seitz (muffin-tin) potential can be used for atoms in the solid state. In recent years, several comprehensive codes calculating the differential cross-sections (DCS) and total cross-sections (TCS) have been published for energies down to very low values (Bote et al., 2009); the database of the total and transport cross-sections is also available (Jablonski et al., 2003).

The appearance of energy losses due to bremsstrahlung radiation is an important process in "elastic" electron scattering by matter. The process of generating continuum radiation, i.e., an energy loss in "elastic" collisions, is due to the change of electron direction connected with photon emission and some additional deceleration. The differential cross-section of photon production with frequency  $\omega$ , after a change in electron direction from  $\mathbf{v}_0$  to  $\mathbf{v}$  (a change in electron direction given by angle  $\gamma$ ) and with angle  $\theta$  between the direction of the incoming electron and photon emission, has been calculated theoretically by several authors (Landau & Lifshic, 1974), (Kirkpatrick & Wiedmann, 1945), (Chapman et al., 1983), (Kissel et al., 1983). Because the cross-sections of bremsstrahlung excitation are relatively very low comparing with other processes, this process is usually omitted.

According to the laws of energy and momentum conservation this means that at electron-nucleus collision some energy is transferred to the nucleus and is thus lost by electron. Due to the big difference between masses of electron and nucleus, the energy transfer to nucleus is usually, e.g., at energies used in electron microscopy and microanalysis, very low, in order of meV. Only at relatively high electron energy (e.g., at TEM) and low atomic number samples (biological samples, polymers etc.) the electron energy loss can be remarkable. According the formula in (Reimer, 1984), the loss at collision of 100 keV electron with carbon nucleus depends on electron scattering angle, at 0.5 and 180 degree being 4.3 meV and 226 eV, respectively.

### 2.2 Inelastic scattering

For simple solution, the simple model of continuous slowing down (CSD) can be employed. In the more exact single scattering calculations, we need to know the inelastic cross sections. The calculation of inelastic differential cross-sections is usually slightly more complicated due to the complicated interaction of charged particle with matter having the various dielectric structure. Moreover, the inelastic cross-section is double differentiated - by change of momentum (depending on scattering angle) and by electron energy loss. For MC simulations we need to calculate the inelastic mean free path (IMFP) and the stopping power for given electron energy and element/material; at inelastic event the result should give us the energy loss and scattering angle.

According to the Bethe theory, the electrons in matter are represented by the system of oscillators and electrons lose energy due to their excitation. If we suppose discrete spectrum  $M$  of oscillators with energy states  $W_i$  and the responsible oscillator strengths  $f_i$ , then we have for stopping power  $S$

$$S = -\frac{dE}{dz} = \frac{\pi e^4}{(4\pi\epsilon_0)^2} \frac{N_1}{E} \sum_{i=1}^M f_i \ln\left(\frac{E}{W_i}\right)^2 = \sigma_0 \frac{N_A}{EA} \sum_{i=1}^M Z_{nl} \ln\left(\frac{E}{E_{nl}}\right)^2. \quad (2)$$

$\sigma_0 = \pi e^4 / (4\pi\epsilon_0)^2 = 6.51 \cdot 10^{-20} \text{ keV}^2 \text{cm}^2$ ,  $N_1$  being the number of atoms in a volume unit ( $N_1 = N_A \rho / A$ ,  $\rho$  the density in  $\text{g/cm}^3$ , or  $N_1 = N_A / A$  if we use mass thickness in  $\mu\text{g/cm}^2$ ),  $N_A$  Avogadro number,  $A$  the atomic mass,  $E$  electron energy in keV and  $z$  the path length in  $\mu\text{g/cm}^2$ . In original Bethe solution, due to the condition  $\sum_{i=1}^M f_i = Z$ , in the some

approximation we can similarly estimate  $f_i$  by electron number at given atomic (sub)shell  $Z_i$  and  $W_i$  by the energy of electrons at these shells by  $E_{nl}$ . This was realized in (Reimer & Stelter, 1986); because of some atomic electrons, especially the outer ones, can result in too large energy losses, the corrected values of  $Z_i$  and  $W_i$  were used. By definition of the mean ionisation potential  $J$  (in eV units) according the condition  $\sum_{i=1}^M f_i \ln W_i = Z \ln J$  we obtain from

(2) the simple formula for so called continuous slowing down approximation of electrons in material

$$S_B = 2\pi e^4 \frac{N_1 Z}{E} \ln\left(\frac{1.166E}{J}\right), \quad (3a)$$

$Z$  being the atomic number and  $J$  the mean ionisation potential, which can depend on  $Z$ . The empirical formula for  $J$  used in the most works on electron probe microanalysis (EPMA), according to (Berger & Seltzer, 1964), is as follows

$$J/Z = 0.015 \text{ for } Z \leq 13 \quad (4a)$$

and

$$J/Z = 9.76 + 58.5 \cdot 0.19 \text{ for } Z > 13. \quad (4b)$$

In practice we can use for the Bethe stopping power formula

$$S_B = 7.85 \times 10^3 \frac{Z\rho}{AE} \ln\left(\frac{1.166E}{J}\right) [\text{eV/nm}]. \quad (3b)$$

The stopping power  $S$  characterizes the energy loss per unit path length (negative derivation of energy dependence on the path length). In reality, the energy losses are not continuous and the "straggling", the large energy losses, may appear as drops in this dependence. The approximation also causes the divergency of  $S$  at decreasing of electron energy to zero. Moreover, the  $S$ , which should increase with decreasing energy, starts to decrease at about  $E = 6.338 J$ . Thus for lower energies the formula for  $S$  can be corrected by empirical corrections - either by the Rao-Sahib & Wittry (RSW) correction (Rao-Sahib & Wittry, 1974) or by the Joy & Luo (JL) correction (Joy & Luo, 1989), also widely used in EPMA calculations. RSW correction is used for energies  $E < 6.338 J$ , then we have for  $S_B$  the formula (the coefficient ensures the continuity of  $S_B$  at the energy  $E = 6.338 J$ )

$$S_B = 6.236 \times 10^3 \frac{Z\rho}{A\sqrt{EJ}} \text{ [eV/nm]}. \quad (5)$$

In the JL correction, instead of mean ionisation potential  $J$  the corrected value  $J'$  is used in equation (3),

$$J' = \frac{J}{1 + kJ/E}, \quad (6)$$

where  $k = 0.77 - 0.85$  for various elements.

Differential cross-section (DCS) of inelastic collision in Born approximation is (Inokuti, 1971)

$$\frac{d^2\sigma}{dWdQ} = \frac{\pi e^4}{E} \frac{1}{WQ} \frac{df(W,Q)}{dW}, \quad (7)$$

where  $E$  is energy of primary electron,  $W$  energy loss,  $Q$  energy given by momentum change ( $Q = \hbar^2 q^2 / 2m$ ),  $\hbar \mathbf{q}$  momentum change,  $\mathbf{q} = \mathbf{k} - \mathbf{k}_0$ ,  $\mathbf{k}_0$  and  $\mathbf{k}$  are wave vectors of impinging and outgoing electrons,  $\mathbf{k}_0 = 2\pi/\lambda_e$ ,  $\lambda_e$  is the electron wavelength and  $df(Q,W)/dW$  is the generalised oscillator strength (GOS) density, which gives the interaction magnitude between electron and target. Integrating the inelastic DCS over all scattering angles, the inelastic scattering is described by energy loss distribution (loss function)  $f(W)$  which can be approximated by relatively simple formula. In (Liljequist, 1978) the function  $f(W)$  is defined as

$$f(W) = \frac{d\sigma_{in}(W)}{dW} \bigg/ \int_{W_{min}}^{W_{max}} \frac{d\sigma_{in}(W)}{dW} dW, \quad (8)$$

where  $d\sigma_{in}(W)/dW$  is the differential inelastic cross section, integrated over the whole solid angle. Then, on the base of supposition of binary collisions of followed electron with electrons of material it was assumed that generally the dependence of  $f(W)$  on energy loss  $W$  is as follows

$$f(W) = HW^{-2} \quad (9)$$

in the interval  $(W_{min}, W_{max})$  and  $f(W)=0$  in  $(0, W_{min})$ ,  $W_{min} > 0$ .  $W_{max}$  and  $W_{min}$  are the limits of electron energy loss in matter. We can use simply  $W_{max} = E$ , the actual energy of an electron,  $W_{min}$  is a constant, which does not change during the electron path.  $H$  is the constant which can be calculated by normalization of  $f(W)$ . If we assume the distribution (9) (hyperbolic), the normalization of  $f(W)$  to unity, i.e.

$$\int_{W_{min}}^E f(W) dW = 1 \quad (10)$$

gives

$$H = \frac{EW_{min}}{E - W_{min}}. \quad (11)$$

In (Liljequist, 1978)), the same shape of  $f(W)$  and the same value of  $W_{min}$  (10 eV) for all the elements was assumed, but better,  $W_{min}$  may be taken as an adjustable parameter. By this way

$$f(W) = \frac{EW_{\min}}{E - W_{\min}} W^{-2} \cong W_{\min} W^{-2}, \quad (12)$$

( $E \gg W_{\min}$ ), and we can calculate the actual energy loss  $W$  at each inelastic event by a way similar to that usually used in the Monte-Carlo codes, by means of the random number  $R$  ( $0 \leq R \leq 1$ ) and the formula

$$\int_{W_{\min}}^W f(W) dW = R. \quad (13)$$

Due to our knowledge of  $f(W)$ , we can also simply calculate  $W_m$ , the mean energy loss,

$$W_m = \int_{W_{\min}}^E W f(W) dW \cong W_{\min} \ln \left( \frac{E}{W_{\min}} \right), \quad (14)$$

knowing  $W_m$  the IMFP (in equations denoted as  $\Lambda_{in}$ ) can be calculated as

$$\Lambda_{in} = \frac{W_m}{S_B}. \quad (15)$$

The advantage of this definition is the presence of only one parameter  $W_{\min}$  for optimizing the agreement of the Monte-Carlo and experimental values, for example the coefficient of elastic reflection  $R_e$ . So, the energy dependence of IMFP is defined by (15) after including (3) and (14), taking into account corrections of  $S_B$  for low energy region. Even though it is not exact shape of loss function  $f(W)$ , the simplicity for fitting of some experimental data is advantageous.

The other possibility is to utilize the supposed analytical shape of energy distribution of losses at inelastic scattering. In the Tougaard theory (Tougaard, 1997), the basic formula for  $f(E, W)$  is

$$f(W) = \frac{BW}{(C + W^2)^2} \quad (16)$$

which is used usually for materials with broad band of energy losses, e.g., Au and Cu (Tougaard, 1997). Here  $W$  is the energy loss and  $B$  and  $C$  are the constants. According to normalisation

$$\int_0^{\infty} f(E, W) dW = \frac{B}{2C} = 1. \quad (17)$$

we have  $B=2C$ . In this formulas,  $f(W)$  does not depend on the primary energy of electron  $E$ . Then from (16) and (17) we can directly write

$$W_m = \int_0^{\infty} W f(E, W) dW = \int_0^{\infty} \frac{BW^2}{(C + W^2)^2} dW = \frac{\pi B}{4\sqrt{C}}. \quad (18)$$

The suggested modification (Starý et al., 2007) can be accomplished by implementation of the energy dependence of the loss function into the Tougaard's formula (16). We carried it out as follows: first, normalisation of the distribution was performed within the limits  $(0, E)$  ( $E$  being the actual energy of electron) instead of  $(0, \infty)$ , which made the former constant  $B$  in equation (19) energy-dependent as follows

$$B = \frac{2C(C + E^2)}{E^2}. \quad (19)$$

Putting equations (18) into (15), and integrate in the limits  $(0, E)$ , the integral can be solved analytically

$$W_m = \int_0^E W f(E, W) dW = \frac{C(C + E^2)}{E^2} \left( \frac{1}{\sqrt{C}} \operatorname{arctg} \frac{E}{\sqrt{C}} - \frac{E}{C + E^2} \right). \quad (20)$$

After calculation of  $W_m$ , we can obtain the proper values of constants  $C$  and using (19) also  $B$  in the analytical shape of energy distribution (16). Then, using the Bethe stopping power  $S_B$  corrected for low energies according to (Joy & Luo, 1989), we calculated the possible values of IMFP with  $C$  as a free parameter for various energies between 0.1 and 20.0 keV. These values were compared with the valid values of IMFP obtained from TPP-2 formula (Tanuma et al., 1991a). The best fit gives the energy dependent value of  $C$  in Tougaard's formula (16) and modified Tougaard's Universal cross section  $f(E, W)$  can be found. The equation (20) can be also solved directly to find the energy dependence of  $C$  using a suitable iteration code. By this way, we obtained for  $C$  the approximate formula  $C = -0.0155 * E^2 + 0.529 * E + 1.8785$  and  $C = -0.0173 * E^2 + 0.6402 * E + 2.1312$  ( $C$ [keV<sup>2</sup>],  $E$ [keV]), in the case of Cu and Au, respectively. Then the energy loss can be calculated according to (13).

At given energy loss, the scattering angle is derived using several suppositions (Raether, 1980):

- scattering angle  $\vartheta_c$  is in interval  $<0, \vartheta_{\max}$ ;
- there is obeyed rule  $\vartheta_{\max} = \sqrt{2\vartheta_p} = \sqrt{W_p / E}$ , energy  $W_{\max}$  is given by  $W_{\max} = \sqrt{4W_p E}$ ;
- if  $W < W_p$ , then scattering angle is  $\vartheta_c = 0$ ;
- if  $W_p < W < W_{\max}$ , then scattering angle is  $\vartheta_c = \sqrt{(W - W_p) / 2\alpha_R E}$ ,  $\alpha_R$  is a constant defined in (Raether, 1980);
- if  $W > W_{\max}$ , the binary collision (electron-electron scattering) takes place and scattering angle is given by formula  $\sin^2 \vartheta_c = W / E$ .

$W_p$  is energy of bulk plasmon (for carbon  $W_p = 25.9$  eV, for copper and gold there is a relatively complicated structure of energy losses); in our code we suppose  $W_p = 30$  eV and 20 eV, respectively;

Except these relatively simple conditions of inelastic scattering the more exact theories appeared. According (Fernandez-Varea et al., 1996) "the calculation of inelastic DCS starts either from the Bethe theory for the inelastic scattering of fast electrons from free atoms or from the dielectric theory for the energy loss of charged particles in condensed matter".

Differential cross-section (DCS) of inelastic collision in Born approximation is given in Bethe theory by (7). Using the GOS density  $df(Q, W)/dW$ , the quantum mechanical Bethe sum rule (instead summing we use now integration)

$$\int_0^{\infty} \frac{df(Q,W)}{dW} dW = Z \quad (21)$$

is obeyed for any  $Q$ . In the limit  $Q \rightarrow 0$  the GOS becomes equal to the optical oscillator strength (OOS) density  $df(0,W)/dW \equiv df(W)/dW$  which describes the excitation of free atom by photons in the dipole approximation.

For binary collisions the scattering angle  $\theta$  is given by formula (Liljequist, 1983)

$$Q = 2E - W - 2\sqrt{E(E-W)} \cos \theta \quad (22)$$

The inelastic mean free path  $\Lambda$  and stopping power  $S = -dE/ds$  are then given as (Liljequist, 1983)

$$\frac{1}{\Lambda} = N_1 \int_Q d\sigma, \quad (23)$$

$$S = N_1 \int_{\Omega} W d\sigma, \quad (23)$$

the mass thickness is used in IMFP and  $S$ , and integration proceeds over space given by kinematic conditions  $\Omega'(W, \Omega)$ . The exact solution of this theory now enables to calculate DCS, IMFP and stopping power for energies of electrons and positrons between 10 eV and 1 GeV (Bote et al., 2009), (Fernandez-Varea et al., 2005). In (Fernandez-Varea et al., 1993), it is possible to find the suitable approximative formulas for calculation of both of IMFP and  $S$ . Some theories use dielectric theory and optical data and calculate the inelastic mean free path and the stopping power by different way, e.g., (Pines, 1964), (Tung et al., 1979), (Ritchie, 1982), (Penn, 1987), (Ding & Shimizu, 1989). The inverse inelastic mean free path is given as

$$\frac{d^2 \Lambda^{-1}}{d\omega dq} = \frac{me^2}{\pi \hbar E} \frac{1}{q} \text{Im} \left( \frac{-1}{\varepsilon(q, \omega)} \right) \quad (24a)$$

where  $q$  is momentum change of electron and  $\omega$  is a frequency giving the energy transfer from electron at inelastic collision,  $W = \hbar \omega$ . Because  $\Lambda = 1/(N_1 \sigma)$  and  $Q = \hbar^2 q^2 / 2m$ , we also have

$$\frac{d^2 \sigma}{dW dQ} = \frac{me^2}{2\pi \hbar^2} \frac{1}{N_1 E} \frac{1}{Q} \text{Im} \left( \frac{-1}{\varepsilon(q, \omega)} \right). \quad (24b)$$

The dielectric response of material with more or less free electron gas and by this way also the IMFP is intensively studied. For free-electron-like material, as e.g. Al and Si, the calculation of  $\text{Im}(-1/\varepsilon(\omega))$  agree very well with the experimental excitation spectrum and contain sharp peak at plasmon energy and several edges due to deep inner shells. Also the theoretical plasmon dispersion relation agrees with experiments, mainly for low  $q$ 's (Raether, 1980). For transition metals and noble metals, the improved method was suggested by (Penn, 1987) and then used in (Tanuma et al., 1987, 1991a, 1991b, 1993, 1994) for calculation of IMFP not only for elements, but also of compounds including organic

compounds. It employs directly the experimental data of  $\varepsilon(\omega)$  for expansion of  $\text{Im}(-1/\varepsilon(\omega, q))$  into infinite series of Drude-Lindhard terms, thus avoiding any fitting parameters. The question of stopping power was solved in (Ding & Shimizu, 1989), where the energy dependences for Al, Si, Ni, Cu, Ag and Au are calculated. The results are different from the Bethe theory for electron energy  $< 10$  keV, for energies 0.1 - 3 keV the differences are substantial.

For practical use, in (Tanuma et al., 1991a) the relatively simple expressions, denoted as TPP-2, are deduced, where the values of IMFP are obtained using material constants for electron energy 0.1 to 2 keV. They are as follows

$$\Lambda_{in} = \frac{E}{E_p^2 [\beta \ln(\gamma E)] - \frac{C}{E} + \frac{D}{E^2}}, [10^{-1} \text{nm}] \quad (25)$$

where  $\beta = -0.0216 + 0.944/(E_p^2 + E_g^2)^{1/2} + 7.39 \cdot 10^{-4} \rho$ ,  $\gamma = 0.191 \rho^{1/2}$ ,  $C = 1.97 - 0.91U$ ,  $D = 53.4 + 20.8U$  a  $U = N_v \rho / M$ ,  $\rho$  is the density [ $\text{g}/\text{cm}^3$ ],  $E_p = 28.8U^{1/2}$  is plasmon energy and  $E_g$  is the energy gap width for semiconductors (in eV),  $N_v$  number of valence electrons and  $M$  atomic or molecular mass. This formula which is usually denoted as TPP-2 gives the possibility to define IMFP of elements, semiconductors and organic materials; sometimes the more exact formula TPP-2M (Tanuma et al., 1993) is utilized.

### 2.3 Surface energy losses

In recent years, the energy losses of an electron transmitted through the surface have been intensively studied. These losses are connected with surface plasmon excitations. The mean number of excited surface plasmons  $P_s$ , which is denoted also as SEP, the surface excitation parameter, can be theoretically given as a function of the electron energy  $E$  and the surface transmission angle  $\vartheta$ , i.e., the angle between the trajectory of the electron and the direction perpendicular to the surface. There are several suggested formulas for SEP, namely (Chen, 1996)

$$P_s = \frac{a}{\sqrt{E} \cos \vartheta}, \quad (26)$$

where  $a$  is the fitting parameter, mentioned in (Chen, 1996) for several materials. Formerly, also another formula

$$P_s = \frac{b}{\sqrt{E} \cos \vartheta} \left[ \left( \frac{\pi}{2} - \vartheta \right) \cos \vartheta + \sin \vartheta \right] \quad (27)$$

was published by the same authors (Chen, 1995). Then, in (Oswald, 1992) for  $P_s$  the other formula was suggested

$$P_s(\vartheta, E) = \frac{1}{a \sqrt{E} \cos \vartheta + 1}, \quad (28)$$

where  $a$  is a parameter which, for nearly free electron gas materials, is given as  $a_{\text{NFE}} = (8a_0/\pi^2 e^2)^{1/2}$  (Werner et al., 2001b). The coefficients  $a/a_{\text{NFE}}$  for elements with other electronic structure can be estimated from the predictive formula in (Werner et al., 2001b).

Usually, a Poisson distribution is assumed for the number of excited surface plasmons. Thus, the probability to excite  $n$  plasmons is

$$P_n = \frac{P_s^n}{n!} \exp(-P_s), \quad (29)$$

and the probability to not excite any plasmon, i.e., to be transmitted without an energy loss, is simply

$$P_0 = \exp(-P_s). \quad (30)$$

Transmission of electrons both into and out of the sample should be taken into account. In the MC code, at each surface transmission the energy and the transmission angle are known. Thus, firstly the SEP is calculated and the random number  $R$  is generated. Then, the normalised accumulated probabilities  $F_n$  of excitations of  $0, 1, \dots, i, \dots$  plasmons were

calculated as  $F_n = \sum_{i=0}^n P_i$ , where  $P_i$  is the probability of generating  $i$  plasmons given by (29),

until its value exceeds the generated random number, i.e.,  $F_n < R < F_{n+1}$ .  $n$  defines the number of excited plasmons and after the electron run in the case, denoted as SEL (see below), their energy is subtracted from the energy of the electron. The abbreviations NEL and SEL will be explained in the next paragraph.

#### 2.4 Computational details and precision estimation

There are several MC codes used frequently for calculation of electron-matter interaction, only some of them are mentioned (Gauvin & l'Esperance, 1992), (Baró et al., 1995). Our MC calculations were realized by our code written in PASCAL language. The conditions of the calculations were set to be identical with the experimental conditions, first of all the solid angle of the detector. Usually, electrons hit to the surface perpendicularly, and the number of electrons reflected or elastically reflected or transmitted into a defined solid angle (defined by detector) were observed. In the case of elastic reflection, the electrons were divided into two groups: NEL and SEL type. Firstly, for elastic reflection, we calculated the electrons which were involved in only elastic collisions, with No Energy Loss during the path in the bulk, and in this case the energy losses at the surface excitations were omitted. The number of electrons was denoted as NEL. The number of electrons which were involved in only elastic collisions and, moreover, without Surface Energy Loss was denoted as SEL; only the electrons without any energy loss including the surface plasmon excitation are taken into account. At general reflection, the energy loss at surface is not subtracted from electron energy for NEL number, and it is subtracted for SEL number.

The hard limits of calculation were set usually to  $10^6$  impinging electrons, which means that the calculation was carried out for approximately 10000 - 100000 (elastically) reflected electrons into a large detector solid angle (RFA) or about 1000 - 10000 into a direction sensitive detector. In spite of low energy correction of  $S_B$  at low energies under 1 keV, its real dependence around 0.1 keV is not well defined. Moreover, when we started our simulations, the lowest energy for which the differential cross-sections of elastic scattering seem to be reasonable was about 0.1 keV for the usually used models (Salvat & Mayol, 1993) (Starý, 1999), (Starý et al., 2004). It can be supposed the decreasing of cut-off electron energy

toward 10 eV should increase calculation time very much. Thus, in our code the electrons which energy decreased under 0.1 keV is taken as absorbed. For these reasons, the minimal energy of the calculations was 0.2 keV for all elements studied. The other details do not differ from those usually used in MC codes. In all the calculations the trajectory of the electron was followed up to the escape from the sample, up the decay of electron energy under 0.1 keV or in the case of following only elastic reflection, up to the first inelastic scattering event, when the single electron run stopped. Then the data for the escaping electron was saved and a new electron started.

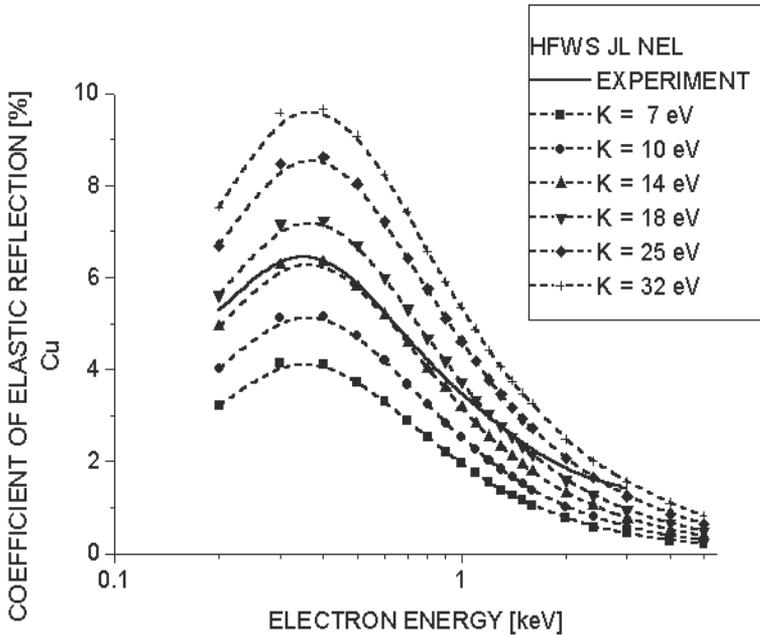


Fig. 1. Definition of optimal value of fitting parameter K for Cu. Comparison with experiment (Schmid et al., 1983).

The reliability of results, e.g., of the IMFP values, is limited by statistical errors. In the case of electron reflection from the bulk material, the number of electrons scattered to the direction sensitive detector is usually between 1500 and 8300 for Al and Au and energies 0.2 and 1.0 keV. The error of the calculated values of the number of elastically reflected electrons (and simultaneously the error of the number of electrons which have lost some energy and/or are scattered into some solid angle, i.e. the number  $N$  of electrons with a given property) can be estimated, if we assume a binomial distribution of the resulting numbers of electrons. In this case, the standard deviation can be estimated as  $s = \sqrt{N_0 \hat{p}(1 - \hat{p})}$ , where  $N_0$  is the full number of electrons and  $\hat{p}$  is the estimated probability of this phenomenon,  $\hat{p} = N/N_0$ . The upper limit of this estimation is  $s_{\max} = \sqrt{N_0/4}$ . At high number  $N_0$  and at low value of probability  $\hat{p}$  this distribution can be approximated by Poisson's distribution, and the standard deviation is approximately  $\sqrt{N}$ . Thus, we can estimate the relative standard deviation of calculations to be between 1% and 3%.

### 3. Monte-Carlo calculations and comparison with experiments

We compared our calculation with experimental values, obtained from literature and/or obtained by cooperation with several experimental laboratories. Electron spectrometers were used for measurements of backscattered intensities. Measurements can also be made using an electron microscope of either transmission or scanning type. The reflection coefficient, the elastic reflection coefficient, the transmission of electrons through a thin foil, the chromatic error and the resolution limit due to energy losses of transmitted electrons and the size of interaction volume in bulk samples and also in thin films have been examined in a big number of experiment, which will be mentioned. The comparison and agreement between measured and calculated values was seeking mainly for the measurable quantities. In the most cases the very good agreement has been obtained.

#### 3.1 Electron backscattering

In our works (Starý, 1999) and (Starý et al., 2004) we simulated the backscattering coefficient  $\eta$  and the coefficient of elastic reflection  $R_e$  of some materials. Firstly, using our MC code, we calculated elastic reflection of electrons. In these calculations the electron path was stopped after inelastic collision and this electron was not taken into the result.

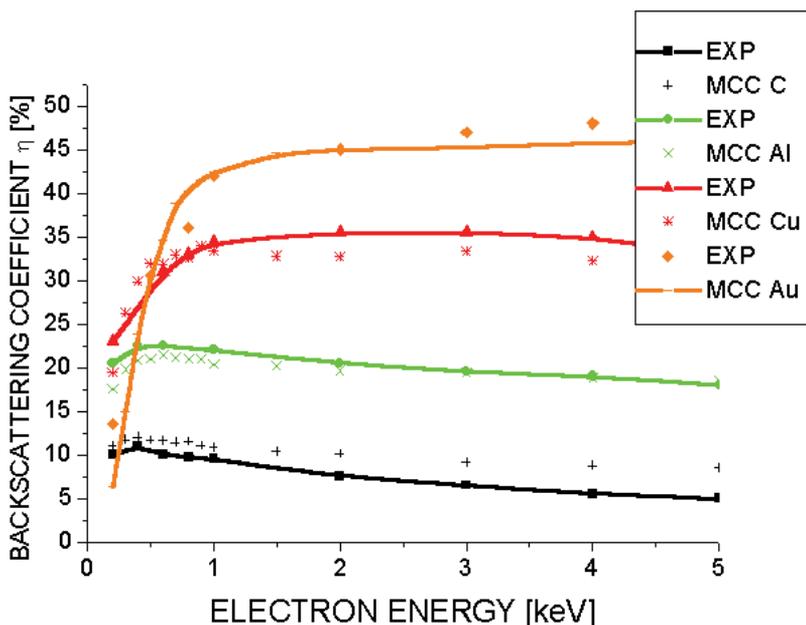


Fig. 2. The comparison of MC calculated energy dependence of backscattering coefficient  $\eta$  for C, Al, Cu and Au with experiment (Fitting, 1974) (lines - measured values, points MC values).

By this way, the energy of followed electrons was the same during the whole path and instead  $W_{min}$ , only parameter K was utilized as the fitting parameter of IMFP. By integration and by comparison with experiment (e.g., with the measurement of elastic reflection) we have found the optimal value of parameter K. The example of this calculation for Cu is in

Figure 1. Optimal value of  $K$  was obtained by interpolation of values of numerical integrals of MC calculated dependences  $R_e(E)$  for various  $K$  and numerical integral of experimental curve between 0.2 and 3 keV (Schmid et al., 1983). In this case, we obtained  $K = 17.17$  eV. Using this empirical parameter given by the best agreement with measurement and equation (3), (14) - where  $W_{\min}$  is replaced by  $K$  - and (15), we are able to find the energy dependence of IMFP. For practical use, e.g., for calculation of energy dependence of coefficient of backscattering  $\eta$ , it is possible to use these values as  $W_{\min}$  in the model similar to (Liljequist, 1978) and calculate the backscattering coefficient. Even though this model is theoretically far from reality, the energy dependence of calculated and measured values of backscattering coefficient  $\eta$  (Fitting, 1974) (Figure 2) shows relatively good agreement. By this way it is possible to use for simulation IMFP values from various sources and compare them. The results of simulation are shown for elastic reflection in the next paragraph.

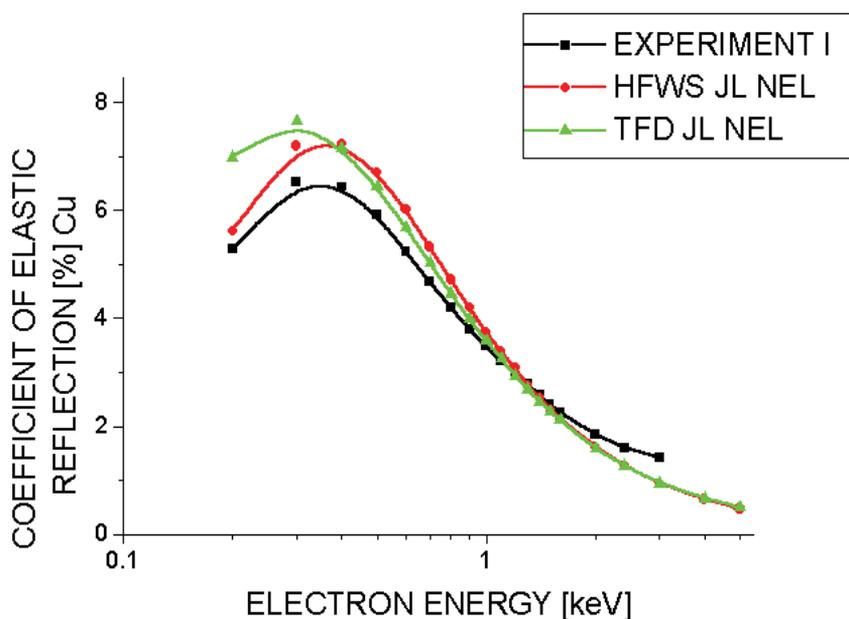


Fig. 3. The comparison of MC calculated energy dependence of coefficient of electron elastic reflection for Cu for two models of elastic DCS with experiment (Schmid et al., 1983).

### 3.2 The influence of elastic and inelastic models on MC results

The next step was comparison of electron elastic reflection from several pure materials using the various physical models of elastic DCS and low energy approximation of the Bethe stopping power, and combinations of these models, to finding the best assessments of simulated and experimental results. The results are shown in Figures 3 and 4. In Figure 3, the values of DCS are calculated by PWADIR code (Salvat & Mayol, 1993) using a static field approximation with relativistic partial wave analysis of the Dirac-Hartree-Fock and Thomas-Fermi-Dirac potential of atoms (DHF and TFD, respectively). Instead of DHF, for atoms in solid state the modified model of atomic potential (muffin-tin) denoted as HFWS model was used. Moreover, two models for improvement of low energy values of Bethe's

stopping power denoted as Joy-Luo (JL) and Rao-Sahib-Wittry (RSW) were compared (Figure 4). To define the agreement of various models of MC simulations and of the experimental curves, an evaluation of the agreement was carried out by calculating the relative differences per one measurement and calculating the residuals (Draper & Smith, 1966) and their root mean squares (RMS) – the differences between measured and calculated values of  $R_e$  in the selected range of energies between 0.2 and 3 keV. The relative differences per one set of measurement were calculated as

$$RD = \frac{100}{n} \sum_{j=1}^n \frac{R_{e,j}^{MC} - R_{e,j}^{exp}}{R_{e,j}^{exp}} [\%], \quad (31)$$

and the root mean squares of residuals were calculated as

$$RMS = \left( \frac{1}{n} \sum_{j=1}^n (R_{e,j}^{exp} - R_{e,j}^{MC})^2 \right)^{1/2}, \quad (32)$$

$n$  being the number of calculated values of energy (18 values). The results are given in Table 1.

ELASTIC MODEL	ELEMENT	C		Al		Cu		Ag		Au	
		RD [%]	RMS								
HFWS	S <sub>B</sub> CORRECTION										
	JL NEL	6.3	0.017	-5.3	0.316	1.8	0.460	41.6	2.018	12.6	0.607
	JL SEL	19.6	0.111	4.5	0.057	1.7	0.306	32.1	1.695	6.3	0.332
	RSW NEL	1.2	<b>0.013</b>	-5.1	0.293	2.9	0.476	33.0	1.787	4.4	0.390
TFD	RSW SEL	14.0	0.104	4.0	0.062	2.4	0.492	24.6	1.500	-0.8	0.224
	JL NEL	7.5	0.037	-6.6	0.377	0.8	0.574	38.1	1.855	19.5	0.672
	JL SEL	24.0	0.121	2.6	0.056	0.6	<b>0.170</b>	29.4	1.565	11.3	0.390
	RSW NEL	3.9	0.028	-6.2	0.355	1.6	0.393	29.8	1.638	8.4	0.319
	RSW SEL	17.4	0.113	1.5	<b>0.048</b>	1.5	0.251	17.2	<b>1.334</b>	1.9	<b>0.160</b>

Table 1. The agreement of measured (Schmid et al., 1983) and calculated values of  $R_e$  for several materials. The combination for each element with best fit is denoted by bold characters.

The set of energies used for RMS calculations prefers the low energy values 0.2 - 1.6 with step 0.1 keV; the next values of energy, which were 2, 2.4 and 3 keV, respectively, have less influence on the RMS value. The obtained values of RD and RMS also indicate that the results for TFD model of elastic scattering are slightly better than those for HFWS model (Table 1 and Figure 3). A comparison of the RSW and JL corrections of the stopping power (Figure 4) shows that the agreement of the calculated and experimental values of  $R_e(E)$  varies for different elements. For C and Al, there is little difference between the RSW and JL corrections, for Cu the JL correction provided better agreement, for Ag and Au better agreement was provided by the RSW correction, though the RMS for Ag is large in comparison with the RMS of the best fit for the other elements. For  $R_e$  values for C, the agreement was approximately the same for JL and RSW corrections, but it was best for the NEL case (this was an exception) and HFWS model.

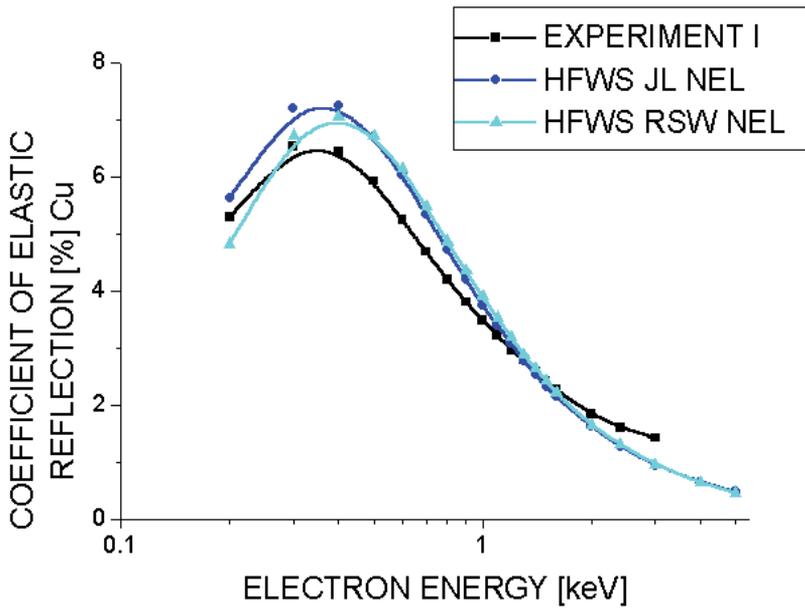


Fig. 4. The comparison of MC calculated energy dependence of coefficient of electron elastic reflection for Cu for two models of Bethe's stopping power for low energies with experiment (Schmid et al., 1983).

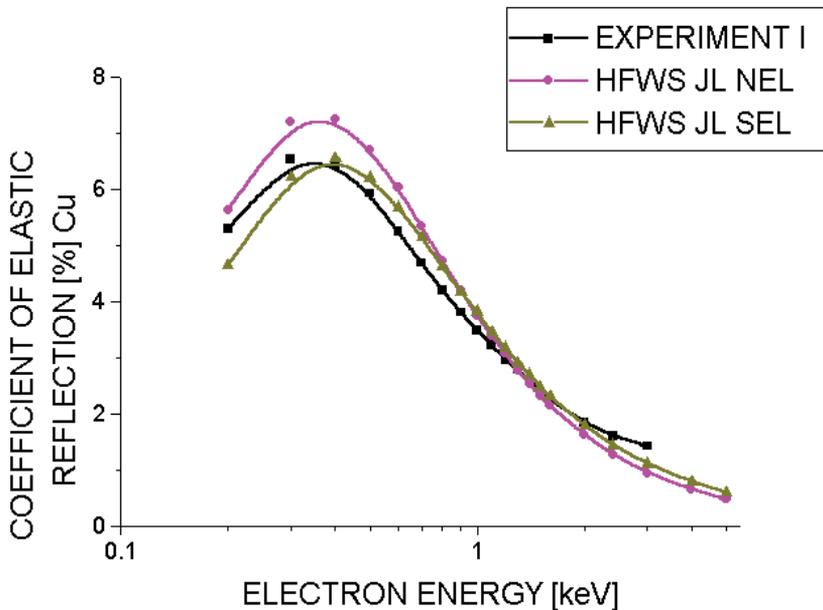


Fig. 5. The comparison of MC calculated energy dependence of coefficient of electron elastic reflection for Cu for two models of presence of surface excitations with experiment (Schmid et al., 1983).

### 3.3 The influence of surface excitations

Also, using the simple model, we compared the results of experimental and MC dependence of electron elastic reflection from several materials on presence of surface excitations. We calculated the energy losses due to the surface transfer in our MC code using the surface excitation probability (SEP)  $P_s$  according to formula (28), the coefficients  $a/a_{\text{NFE}}$  for Al, Cu and Au were 0.7, 2.0 and 1.5, respectively (Werner et al., 2001b); for C and Ag the  $a/a_{\text{NFE}}$  coefficient was assumed to be 1.0 and 2.0, respectively. Transmission of electrons both into and out of the sample was taken into account. We calculated in each direction both the number of electrons reflected in this direction with No Energy Loss (NEL) in bulk and the number of electrons both without energy loss in bulk and without Surface Energy Loss (SEL).

Even a visual comparison of the results obtained without and with surface excitations into account (NEL and SEL, respectively) clearly shows that incorporating the surface excitations (SEL model) improves the agreement of the measured and calculated values of  $R_e$  for the HFWS model of elastic collisions for Al, Cu and Au (e.g., for Cu Figure 5). The preference of SEL values is clearly shown in table 1 for Al, Cu, Au and Ag.

Qualitatively, the influence of taking surface energy losses into account appears as a tilting of the MC curves around the pivot point near 1 keV. This results from the fitting method, where the area of the whole curve is fitted, but the influence of the surface excitations is more intensive for low energies.

### 3.4 Inelastic mean free path definition

Finding the optimal value of fitting parameter  $K$  and using value of  $S_B$ , also IMFP can be found. The MC calculated IMFP values and their dependence on energy in the observed energy range for Cu and Au are in Figures 6.a,b (TPP-2 denotes values of IMFP calculated using formula from (Tanuma et al., 1991a), FVL means values of IMFP calculated using formula in (Fernandez-Varea et al., 1993), HFWS, JL means type of elastic scattering model and low energy corrections of  $S_B$ , and NEL and SEL are the models of surface excitation. It is seen that differences between the IMFPs calculated by TPP-2 formula or formula from (Fernandez-Varea et al., 1993) are minimal. In spite of supposition, the values of IMFP found by comparison of MC and experiment using the fitting procedure are better for the NEL model than the SEL model of surface excitation. The NEL values of IMFP obtained without incorporating the surface energy losses into the calculation agree reasonably well with the theoretical values of TPP-2 (Tanuma et al., 1991a) and also those in (Fernandez-Varea et al., 1993), and for all the elements the difference lies within the range of ~10% .

The shown MC deduced IMFPs were obtained from the optimal combination of the two corrections of stopping power (RSW and JL), and for the two cases of incorporating the surface energy losses (NEL and SEL). For the resulting IMFPs the differences between the two models of elastic collisions (TFD and HFWS) are very small. For Al and Cu, the agreement of IMFP for HFWS NEL case with the theoretical values of TPP-2 is the best; for C and Cu the MC calculated values are slightly lower, and for Al, Ag and Au they are slightly higher than the theoretical values of TPP-2. The MC - NEL defined values of IMFP agree in the energy range 0.2-1.5 keV also with the IMFP values in (Ashley, 1988).

For the assessment, we used our simulated data of  $R_e$  for the set of fitting parameters to calculate IMFP from experiment (Schmid et al., 1983) for several single values of energy by the usual EPES method, employed also in (Dolinski et al, 1988a, 1988b and 1992). We

obtained the values of IMFP, very similar to those used in our MC code for generating the  $R_e(E)$  with the best fit. Moreover, exchanging the experimental data of  $R_e$  for the data calculated by our code, we obtained using the EPES process exactly the IMFP used for generating  $R_e$ .

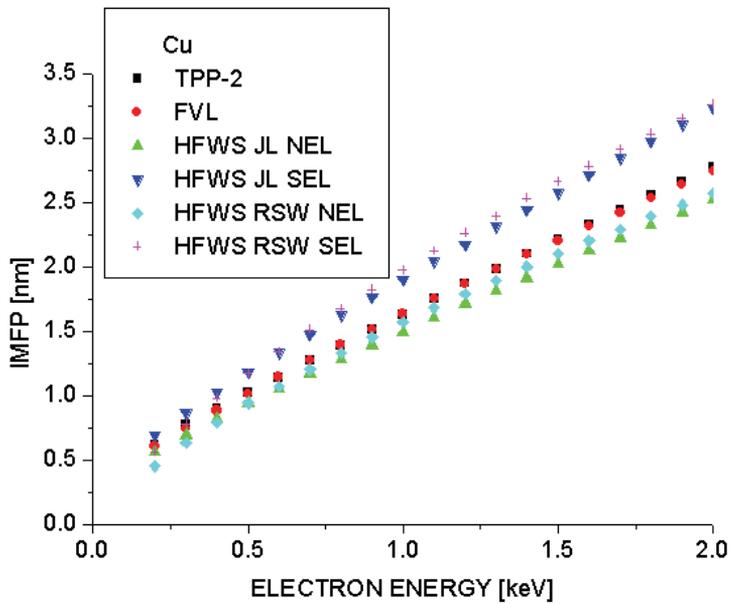


Fig. 6.a The dependence of IMFP on electron energy for Cu.

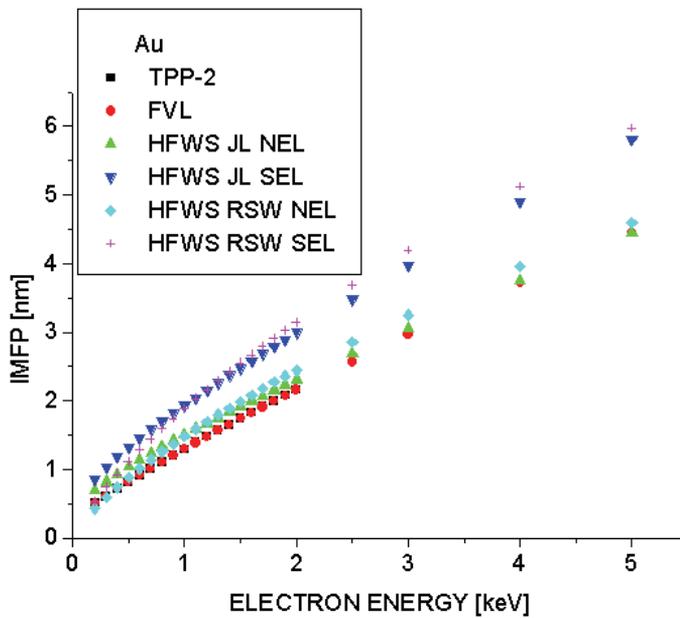


Fig. 6.b The dependence of IMFP on electron energy for Au.

When we evaluate the RFA data at single values of energy, the IMFP values deduced from MC calculations start to increase at 1.5 keV more intensively than is expected according to the TPP-2 formula. This phenomenon did not appear in our IMFP values calculated using a numerical integral. We suppose that this again results from inaccuracies of  $R_e$  measurement by RFA, which is minimised by our fitting method, inasmuch this inaccuracy has a slight influence on our IMFP values.

In evaluating the experimental data, the procedure described as „cleaning“ the experimental data from surface energy losses is sometimes used (Werner et al., 2003) and (Jung et al., 2003). Unfortunately, this procedure works only if the detector is able to resolve the angular distribution of the elastically reflected electrons. For measurements using an RFA energy analyser, where the integral values of  $R_e$  are measured, „cleaning“ cannot be carried out, because we would need to know the angular dependence of electron reflection in the experiment, i.e., the absolute  $R_{e,9}$  values. In our code, „cleaning“ can be performed by fitting the  $R_e$  values by the SEL procedure, definition of optimal IMFP for it and then in NEL case of this simulation to obtain the  $R_e$  values that correspond to the „cleaned“ experimental data.

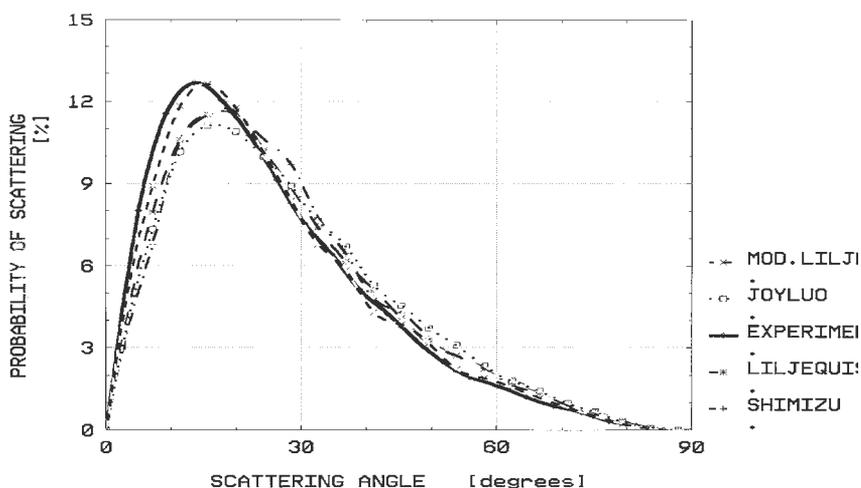


Fig. 7. Comparison of experimental and calculated values of angular distribution of transmitted electrons, and several models of inelastic scattering. Material Al,  $E_0 = 20$  keV, film thickness  $94.9 \mu\text{g}/\text{cm}^2$ , i.e., 351 nm. Models: MOD.LILJI=G, JOYLUO=D, LILJEQUI=F, SHIMIZU=E.

### 3.5 Electron transmission through the thin samples

For the oldest simulation where the DCS was given usually by analytical formulas the simplest models were used:

There was for elastic differential cross-sections

- A. Rutherford's model with screened potential of nucleus (Reimer, 1984),
- B. Mott's model including electron spin (according (Reimer & Lodding, 1984)),
- C. tables, calculated for Hartree-Fock model of atomic potential (Riley et al., 1975),

and for inelastic differential cross-section and IMFP

- D. Bethe's continuous slowing down (CSD, at low energies using empirical correction of  $S_B$  (Joy & Luo, 1989)),
  - E. individual scattering using exponential distribution of energy losses (Shimizu, 1975),
  - F. individual scattering using hyperbolic distribution of energy losses (Liljequist, 1978),
- In both E and F models the direction change comes to depend on the energy loss;
- G. model, similar to F, without directional change at inelastic scattering.

We will show some comparison with experiments because also very simple models gave relatively reasonable results, even though the electron number for simulation was relatively low.

We simulated the transport of electrons of energy 10, 20, 50 and 100 keV through the thin film of C, Al, Cu and Au with thickness 20-10000 nm (Starý, 1996). The calculated data were compared with experimental values (Cosslett & Thomas, 1964a, 1964b, 1965), (Reimer et al., 1978); we evaluated

- electron backscattering coefficient,
- angular distribution of transmitted electrons,
- energy distribution of transmitted electrons.

The example of comparison of angular and energy distribution of transmitted electrons for Al and Au, respectively, and energy of primary electrons  $E_0 = 20$  keV is in Figures 7 and 8. This comparison has two disadvantages:

- it is valid only for single material, one value of electron energy and one film thickness,
- the visual comparison is not exact.

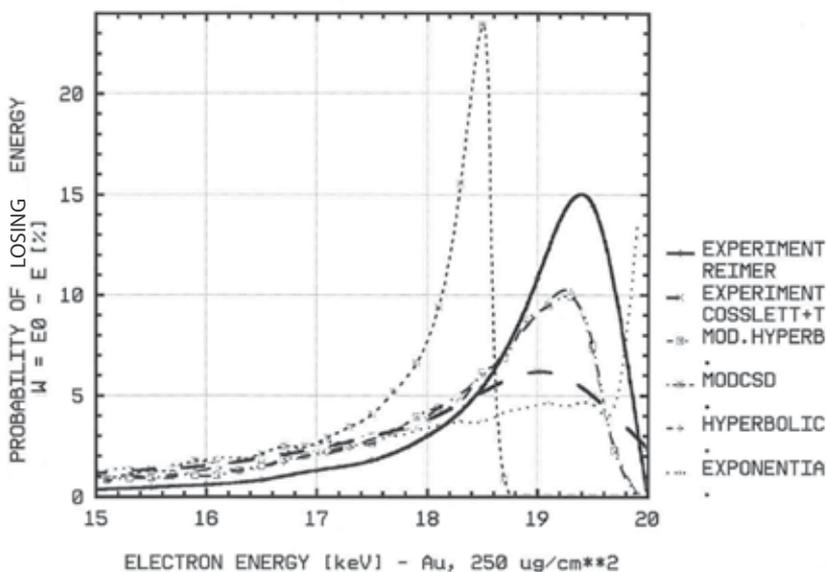


Fig. 8. The dependence of probability of energy loss  $W$  on energy of transmitted electrons for several models of inelastic scattering. Material Au,  $E_0 = 20$  keV, film thickness  $250 \mu\text{g}/\text{cm}^2$ . Models: MOD.HYPERB  $\equiv$  G, MODCSD  $\equiv$  D, HYPERBOLIC  $\equiv$  F, EXPONENTIAL  $\equiv$  E.

We were therefore seeking a way to compare results simultaneously for some range of conditions (energy, thickness). Instead of coefficient of reflection  $\eta$  we are using the Niedrig's formula  $C_N = \eta / (N_1 Z^2 d_{1/2})$ ,  $Z$  being the atomic number,  $N_1$  number of atoms in

volume unit and  $d_{1/2}$  the thickness of the film when the backscattering coefficient is equal to the half of the backscattering coefficient value on bulk material (Niedrig, 1982). The results of all the materials can be theoretically comprimed in one graph (excluding carbon for lack of experimental data).

For other comparison we can use residual method using squares of residuum  $R_s$ . The number of experimental results was 32 (Al, Au) and 10 (Cu). Thus we fitted 4 simulated values (for energies 10, 20, 50 and 100 keV) by function  $C_N = A E_0^{-B}$  and experimental values were compared with  $C_N$  values calculated using fitting parameters A, B and value of energy  $E_i$  in experiments. Then

$$R_s = \frac{1}{n} \sum_{i=1}^n (C_{i,exp} - A E_0^{-B}). \tag{33}$$

In the case of simulation with the same parameters (thickness, energy) as experiment, and if the number of simulations and measurements was the same, we can use other formula

$$R_s = \frac{1}{n} \sum_{i=1}^n (X_{i,exp} - X_{i,calc})^2, \tag{34}$$

where  $X_{i,exp}$  a  $X_{i,calc}$  are the experimental and simulated values of various quantities. Lower value of  $R_s$  means better agreement with experiment. The results are in tables 2 and 3.

Model of elastic scattering	Element								
	Al			Cu			Au		
	A	B	$R_s$	A	B	$R_s$	A	B	$R_s$
Experiment	12913	1.922		2035	1.392		2439	1.549	
Rutherford's model	8629	1.880	839	10524	1.906	23	9513	1.847	254
Mott-Reimer's model	10598	1.947	722	6486	1.754	26	2871	1.522	30
Tables - Hartree-Fock	8860	1.816	464	6369	1.674	42	3180	1.555	12

Table 2. The comparison of experimental and calculated values of coefficients A and B (coefficients of fitting function  $C_N = A E^{-B}$ ), and  $R_s$  (sum of residuum squeres) for several models of elastic scattering for Al, Cu and Au.

Model of inelastic scattering	Element		
	Al	Cu	Au
Shimizu m.	16.4	9.6	18.5
Liljequist m.	7.9	2.1	15.0
Modified Liljequist m.	4.3	1.9	5.2
CSD + Joy-Luo m.	13.3	8.9	27.3

Table 3. The  $R_s$  values from comparison of experimental and calculated values of the most probable scattering angle of transmitted electrons for Al, Cu and Au and primary energy 20 keV.

In electron microscopy and microanalysis, the MC simulation can be used to find the values of some quantities whose experimental value can be measured only with some complications. Thus, we decided to apply the MC method for low-voltage transmission

electron microscopy (LVEM). The LVEM was recently developed for observing unstained biological samples and organic thin films, and is able to work in transmission, scanning transmission and reflection modes (DeLong et al., 1998). The main advantage of this method consists in the use of an accelerating voltage around 5 kV with final light microscope magnification of an electron-microscopical image. It delivers nearly twenty times more image contrast enhancement than a high voltage electron microscope using accelerating voltage 100 kV (Lednicky et al., 2000). In the case of biological specimens, staining procedures can therefore be omitted. On the other hand, such lowly-accelerated primary electrons are able to pass only through ultra thin sections, below 20 nm in thickness. Electrons accelerated by energy 5 keV are able to scatter intensively on atoms with a low atomic number including atoms of resins, which also participate in scattering and contribute to image formation. This means that the microscope is very sensitive to specimen thickness.

The aim was to show the relation between image contrast and specimen thickness, which is important for defining the sample thickness directly from electron microscopical observation, and to estimate the critical thicknesses for elastic and inelastic scattering in a sample, which characterize the possibility to observe biological samples under given conditions.

In the theory of image contrast formation in an amorphous material, an absorption contrast is assumed, which depends on the thickness of the sample and on its atomic number and mass (Reimer, 1984). The logarithmic contrast in electron microscopy is usually defined as  $C_0 = \log N_0/N$ , where  $N_0$  and  $N$  are numbers of electrons (intensities) incident to the sample and those transferred through the objective aperture after sample transmission, respectively. The contrast of the sample can be characterized by the mass-thickness contrast parameter  $S = N_A \sigma_t/A$ ,  $N_A$  is the Avogadro number,  $\sigma_t$  is the total scattering cross-section of atoms in the specimen, and  $A$  is atomic mass. Then, we can write for  $C_0 = S\rho x$ ,  $\rho$  being the sample density,  $x$  the geometrical thickness of the specimen and  $\Delta C_0 = S(\Delta\rho)x$ ,  $\Delta C_0$  and  $\Delta\rho$  being the contrast and density differences at given thickness, respectively. Linear dependence of the contrast on the film thickness can be assumed for low density materials up to a certain critical thickness, given by the onset of multiple scattering. The characteristic "critical thickness" is defined as the thickness at which the electrons proceed on an average one elastic or inelastic scattering event during sample transmission. Next, the image is also deteriorated by inelastic scattering of the electrons, which is non-localized and decreases the image resolution.

Here, the elastic DCSs were calculated by the ELSEPA code (Salvat et al., 2005), using Dirac-Hartree-Fock with the muffin-tin atomic model. To simulate the IMFP we used values given by the TPP-2 formula (Tanuma et al., 1991a) or those obtained from our MC evaluation of absolute Elastic Peak Electron Spectroscopy.

For calculating the energy loss distribution we used the modified Tougaard's model of the Universal cross-section of electron inelastic scattering  $f(E,W)$  (Starý et al., 2007), so we were able to find the values of  $C$  for several electron energies between 0.1 and 5.0 keV. Then we found the functional dependence of  $C$  on electron energies by regression. The approximate formula is  $C = 571.64 \ln(E) - 2615.9$  ( $C$ [eV<sup>2</sup>],  $E$ [eV]). The results obtained by calculating  $C$  using TPP-2 values of IMFP were slightly different. The former method was used in our calculations. Energy loss functions calculated in this way have a reasonable shape with a maximum at about 20 - 30 eV. The amount of energy loss in an inelastic collision was obtained from the energy loss distribution by the standard Monte Carlo procedure. Then, the scattering angle was calculated either for plasmon scattering, when the scattering angle  $\vartheta_c$  is between a minimum value  $\vartheta_p$  (Reichelt & Engel, 1984)

$$g_p = \frac{W_p(E + E_0)}{E^2 + 2EE_0}, \tag{35}$$

and a maximum value  $g_{\max} = \sqrt{2g_p}$  or by the second type of scattering, electron-electron scattering with larger energy loss ( $W \geq 100$  eV); the scattering angles  $\vartheta_c$  are also larger and they are given by the formula (Reimer, 1996)

$$\sin^2\vartheta_c = W/E \tag{36}$$

The energy of bulk plasmon is  $W_p=25.9$  eV (Raether, 1980).

The LVEM 5 electron microscope has two objective apertures with diameter 30 and 50  $\mu\text{m}$  (angular apertures 0.68 and 1.14 deg, respectively). The energy for the simulation was 5 keV, and the values of the apertures sizes for calculation were taken as 0.62 and 1.2 deg, respectively. We studied pure carbon with density between values  $\rho = 1.60 - 2.34$  g/cm<sup>3</sup>. The hard limits of the calculation were set to 1 000 000 primary electrons.

The MC results on electron transmission show an approximately exponentially decreased number of electrons without scattering with increasing of sample thickness. The different geometrical thicknesses are given by three values of carbon density (Figure 9). Figure 10 shows the change in logarithmic contrast C for electrons with output angle  $\vartheta \leq 11$  mrad with thickness. A very low degree of nonlinearity appeared for these dependences up to 60 nm. The dependences are very similar for the two models of scattering angle distribution for inelastic scattering.

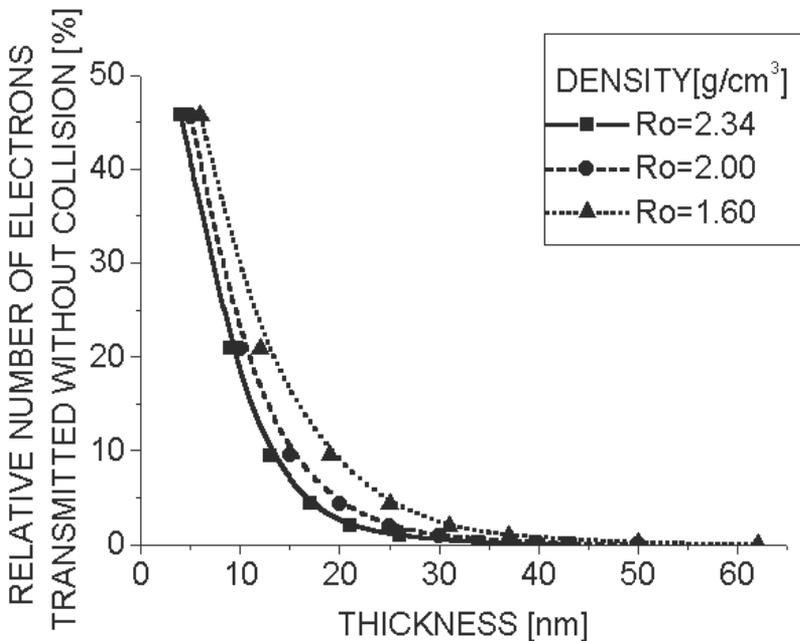


Fig. 9. Dependence of electron transmission through the film on film thickness.

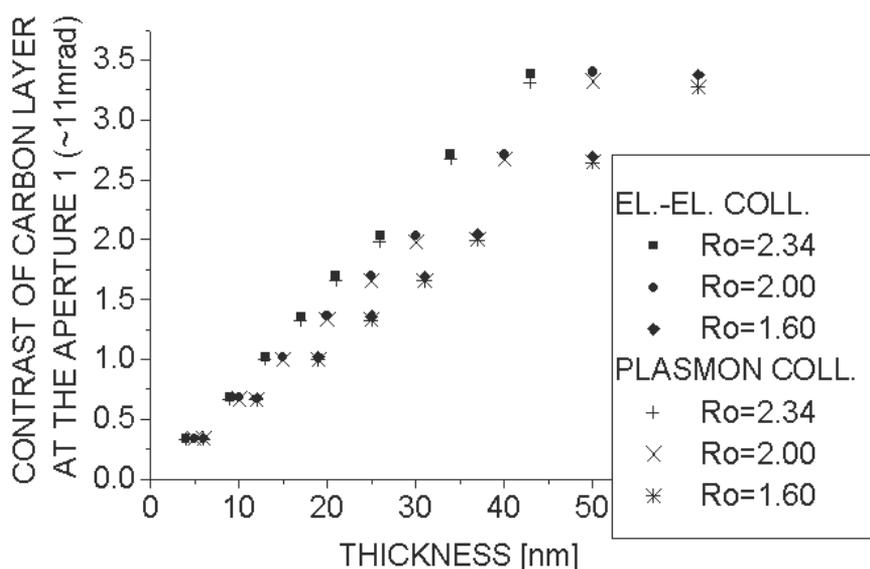


Fig. 10. Dependence of logarithmic contrast on film thickness.

DENSITY [g/cm**3]	REGRESSION EQUATION FOR ELASTIC COLLISIONS	TRANSPARENT THICKNESS [nm]		
		ELASTIC	INELASTIC	ALL
2.34	$y = 0.0001x^2 + 0.0614x + 0.0478$	15.0	8.9	5.5
2.00	$y = 0.0001x^2 + 0.051x + 0.0683$	17.7	10.3	6.3
1.60	$y = 2E-05x^2 + 0.0444x + 0.039$	21.4	12.8	7.9

Table 4. Critical transparent thickness calculated for elastic, inelastic and all collisions.

The characteristic "critical thicknesses" were obtained from the dependence on thickness of the relative collision number for electrons transmitted through the film (Table 4). Because the density of a real carbon film is assumed to be higher than the density of the biological samples (about 1.6 g/cm<sup>3</sup>), the calculated values should be the limits of microscopic observations in LVEM.

The results showing the influence of objective aperture size are shown in Figure 11. The logarithmic contrast  $C$  for all thicknesses scarcely depends on the objective aperture up to tens of milliradians, then the contrast starts to decrease. This is due to high intensity in the zero collision peak, because even at this objective aperture size the number of scattered electrons is comparable with the zero collision peak. Again, this result is the same for both models of angular distribution of inelastically scattered electrons. Finally, Figure 12 shows the dependence of the cumulative electron number (not including the zero collision peak) of electrons transmitted through the film into the aperture. Only here did a difference appear between the models of the angular distribution of inelastically scattered electrons. The calculated thickness and the angular dependences of the contrast in Figures 9 and 10 agree with the assumed values and shape of these dependencies at the energy and in the thickness range used here. The dependences in Figure 10 show the limits of the increase in contrast due to decreasing objective aperture size.

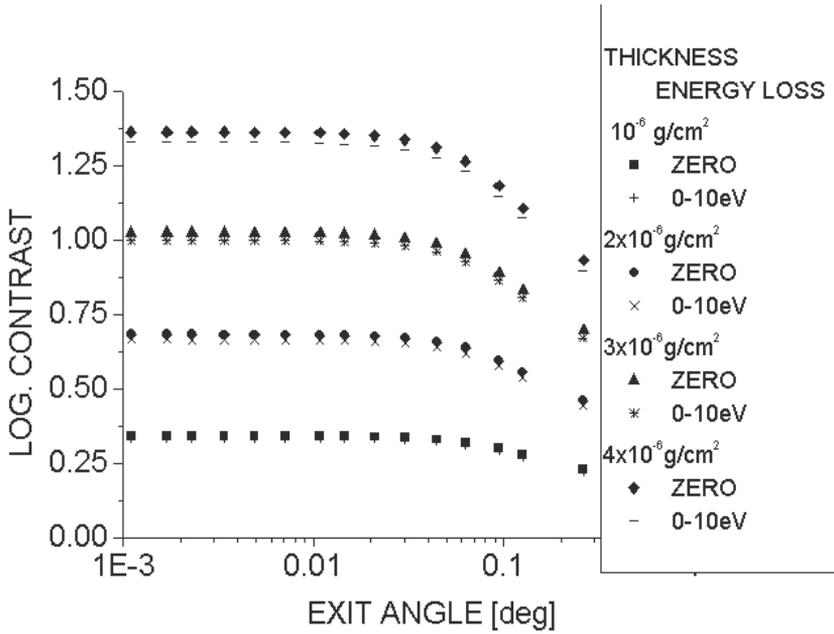


Fig. 11. Dependence of logarithmic contrast on the maximal output angle (i.e., the size of objective aperture) for electrons with ZERO energy loss and with energy loss 0-10 eV, at several values of film thickness (in g/cm<sup>2</sup>).

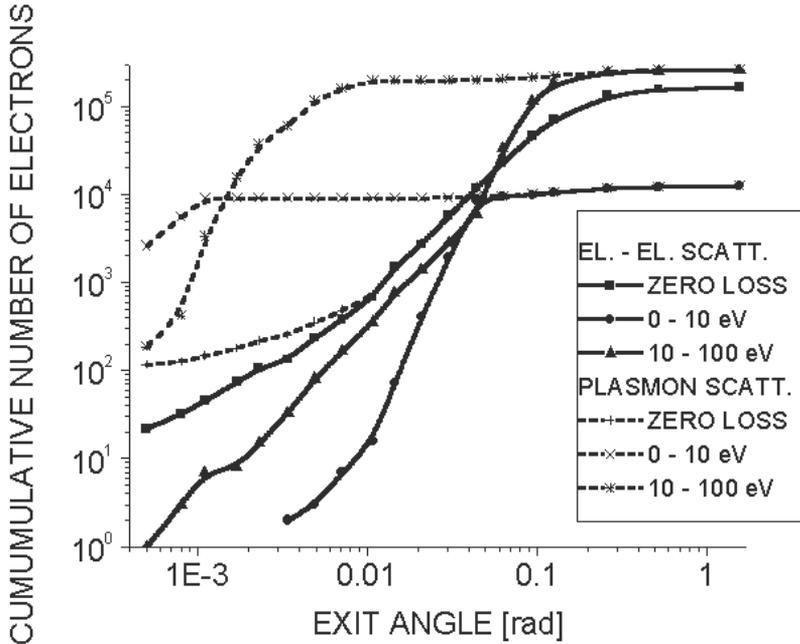


Fig. 12. Dependence of cumulative electron number of electrons transmitted through the film on objective aperture (thickness 10<sup>-6</sup> g/cm<sup>2</sup>).

### 3.6 Interaction volume

Among the most important parameters are the dimensions of the interaction volume, i.e., the dimensions of the volume at which the interaction of electrons with matter produces X-ray radiation. This quantity is important mainly when evaluating the composition of non-homogeneous samples, e.g., thin films on a substrate or particles in a matrix. Usually, the diameter of interaction volume  $D_{90}$  is defined as the diameter containing 90% of the produced (emitted) X-ray intensity. Simultaneously, the depth containing 90% of the produced (emitted) X-ray intensity can be defined as the information depth. Another definition of information depth and, simultaneously, a possible way of calculating it, is given by the decay of electron energy under the lowest energy necessary for excitation of a given X-ray radiation. This depth can be measured or estimated for example using the Bethe formula for stopping power. The actual diameter of the interaction volume can be experimentally measured by moving an electron beam across the sharp boundary of two materials of different composition and simultaneously detecting the signal from element contained only in one material. However, the interface sharpness must be guaranteed and interdiffusion must be avoided, which is not too simple. MC simulation can directly provide the distribution of X-ray production in matter (Murata et al., 1987) and (Kyser, 1989), and in this way can establish some limits of the dimensions of the interaction volume. The dimensions of the interaction volume should be compared with experimental estimations, but these values can be measured only indirectly and with relatively low precision; only the more or less precise semi-empirical formulas that are usually used provide only a rough estimation.

In our previous work (Starý, 2003) we have tried to define the dimensions of the interaction volume not only in homogeneous bulk samples, but also in case of film on the substrate. Moreover, we also tried to calculate their dependences on the electron energy and the thickness of the film. In this work, we used for calculation of the elastic DCS the PWADIR code with a muffin-tin model of atomic potential (Salvat & Mayol, 1993), the IMFP is obtained using optimization of the values of  $W_{\min}$  (Starý, 2000), with  $W_{\min} = 18.93$  eV for Au and  $W_{\min} = 9.52$  eV for Si, so the obtained values of IMFP agree well with theoretical values given in (Tanuma et al., 1991a). Also the corrected Bethe stopping power was used. The surface plasmon excitation at input and output of the electron in the sample and also the energy losses at elastic collisions according to (Starý & Jurek, 2002) were taken into account. We simulated the process for a thin Au film on a thick Si substrate (i.e., not transmittable for electrons of the energies used). The electron beam energy was in the range 5 - 30 keV, the film thickness was in the range 0.05 - 1.0  $\mu\text{m}$ . The number of primary electrons was relatively low (2000), but the number of inelastic collisions was  $10^5$  -  $10^6$ . As the backscattering coefficient had a relative standard deviation of about 2%, we suppose the number of trajectories is reasonable for relatively reliable results. In the calculations, we copied the conditions of measurement, especially the take-off angle of the X-ray detector (TOA =  $40^\circ$ ).

In the code, MC simulation of the electron trajectory in a stratified sample uses the algorithm for transmission of electrons into the next layer with different composition (Murata et al., 1987), with some corrections. In this model, the accumulated probability of collision increases during the electron path, and after transmission through the interface an electron can employ only the remnant of the accumulated probability; also, after transmission the electron keeps its direction. Also X-ray intensities were calculated in this work from the probability of photon emission at an inelastic collision. For characteristic

photon excitation, the Powell cross-sections (Powell, 1989) with the Schwaab coefficients (Schwaab, 1987) were used

$$\sigma_{c,P} = \sigma_0 \frac{b_j Z_j}{E E_c} \ln \left( c_j \frac{E}{E_c} \right), \quad (37)$$

where  $E$  is the energy of an electron,  $E_c$  is the critical energy for ionisation of the shell in question,  $Z_j$  is the number of electrons in the  $j$ -th shell, and the constant  $\sigma_0 = 6.51 \times 10^{-20} \text{ keV}^2 \text{ cm}^2$ .

Because we have one element in the substrate and one element in the film, the actual element type is selected according to the position of electron (being either in film or in substrate). The ionisation probability is calculated at each inelastic scattering collision and at a given depth of the actual collision in sample  $z$  with actual electron energy. The intensity  $I_c$  of the element in question (i.e., the total number of photons of characteristic radiation of the element, detected in a solid angle  $\Delta\Omega$  of a hypothetical detector, and generated by the given number of electrons) is then calculated as

$$I_c = \frac{1}{4\pi} N_1 \sigma_c \omega q t, \quad (38)$$

where  $\sigma_c = \sigma_{c,P}$  is defined above,  $N_1 = N_A \rho / A$ ,  $\omega$  is the fluorescence yield calculated according to (Burhop, 1955),  $q$  is the ratio of the intensity of a measured line to the intensity of the whole X-ray line family (e.g.,  $I_{K\alpha} / (I_{K\alpha} + I_{K\beta})$ ), and  $t$  is the actual trajectory before collision.  $I_c$  is summed during the whole run of program. The intensity of the emitted radiation is calculated from the produced intensity, supposing the exponential decay of this intensity given by the mass absorption coefficients of the film and substrate (Goldstein, 1981) and taking into account the actual take-off angle of the detector.

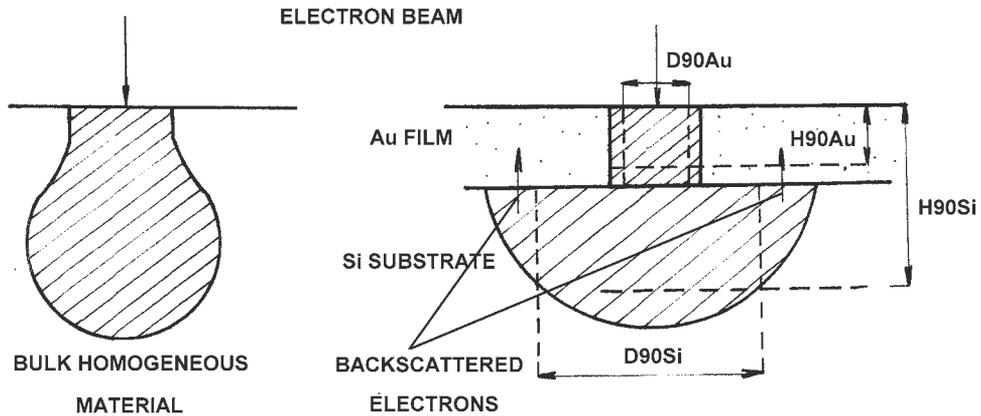


Fig. 13. The spectacular shape of the interaction volume, and definitions of the diameter of interaction volume  $D_{90}$  and the bottom depth of interaction volume  $H_{90}$ .

We simulated the dependences of the dimensions of the interaction volume on the electron energy and film thickness. The volume distribution of the produced X-ray intensities gives the interaction volume in the sample. The spectacular shape of the interaction volume and

the definitions of the diameter of interaction volume  $D_{90}$  and the bottom depth of interaction volume  $H_{90}$  are shown in Figure 13. This Figure was drawn according to a qualitative image of the interaction volume, using the simple MC code from (Ly & Howitt, 1992). The interaction volume in the substrate in the vertical direction starts at the film/substrate interface and continues down into the substrate. In our code, the cell dimensions are defined in mass thickness units ( $\text{g}/\text{cm}^2$ ) - in these units, the size for materials of different density is the same; the real size (e.g., in nanometers) is different.

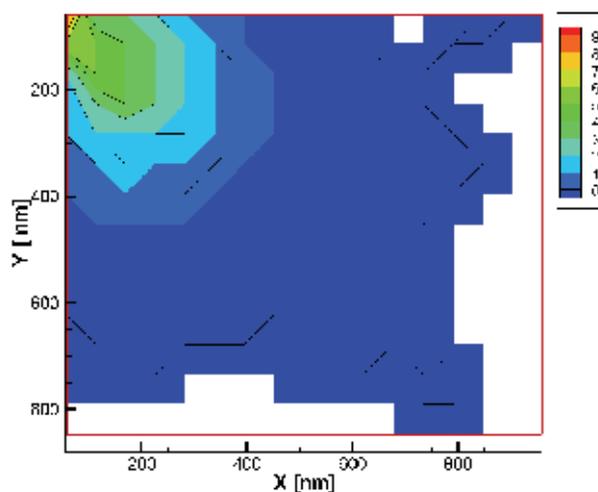


Fig. 14.a The interaction volume in bulk Au sample;  $E_0 = 20$  keV. The intensities are integrated around the Y axis of image, electron beam of zero diameter hit the sample perpendicular to the surface at left upper corner.

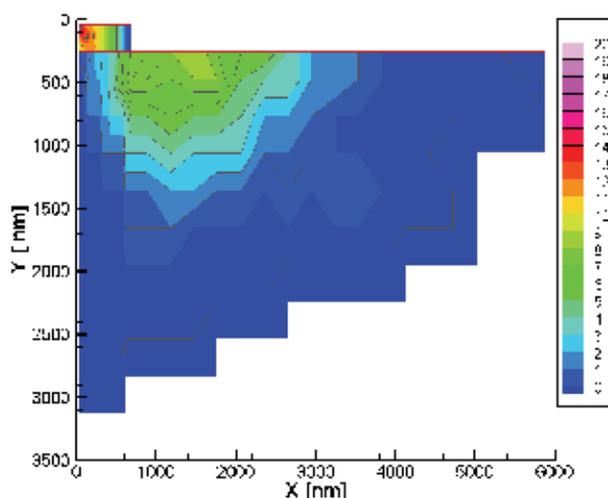


Fig. 14.b The interaction volume in the sample composed from thin film of Au on bulk Si sample;  $E_0 = 20$  keV, film thickness  $\sim 250$  nm. The same conditions of electron excitation, the different scale at the axis.

The Bethe range is the maximum supposed depth of electron interactions in the sample, and the maximum depth of X-ray production is limited by the energy necessary for X-ray excitation. Thus, firstly, the Bethe range and the maximum production depth in a layered structure composed of several elements are calculated. If the maximum production depth is lower than the whole sample thickness, its value is divided into 20 depth divisions. Otherwise, the sample thickness is divided in the same way. Next, radial division into 20 divisions of the same size is prepared. The X-ray produced radiation is placed into these cells, using the position of the actual inelastic collision; the intensities are calculated from the probabilities of X-ray radiation excitations. The intensities into radial cells are summed from the whole volume in a hollow cylinder of given dimensions, i.e., the signal is integrated over all azimuthal angles. The results for bulk Au target and for Si bulk sample with Au thin film are in Figures 14 a,b.

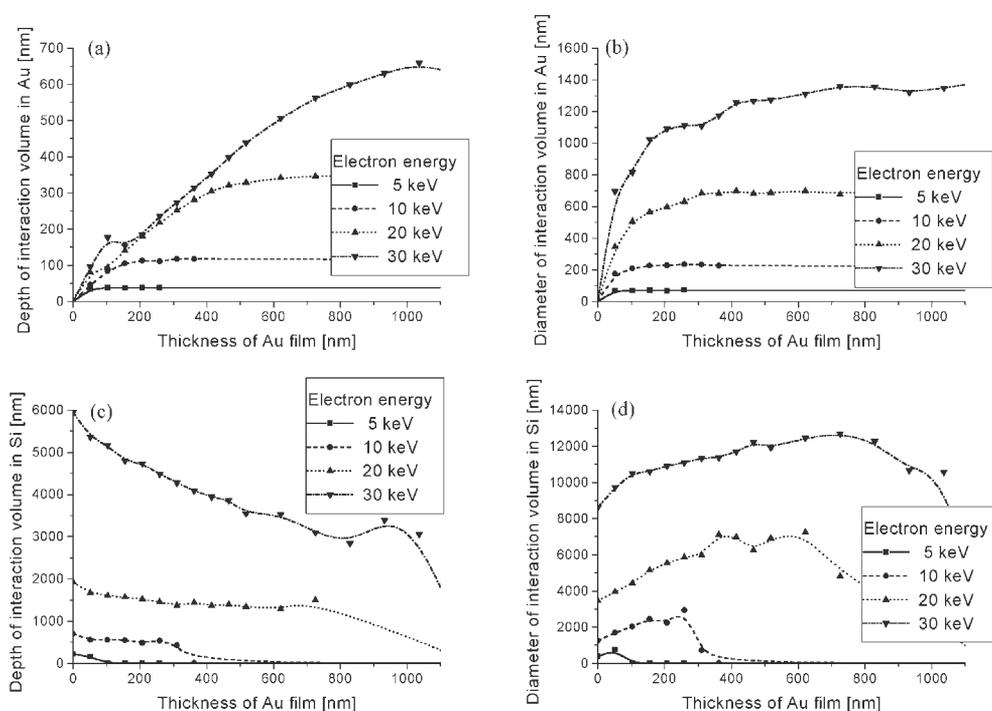


Fig. 15. Dependence of the depth and radius of the interaction volume on electron energy and film thickness: (a) depth of the interaction volume in the Au film (H90Au); (b) diameter of the interaction volume in the Au film (D90Au); (c) depth of the interaction volume in the Si film (H90Si); (d) diameter of the interaction volume in the Si film (D90Si).

Both radial and depth profiles of X-ray production in the samples were also calculated. Figure 15 shows the dependences of the depth and radius of the interaction volume on film thickness. Figure 15a shows that the depth of the interaction volume in the film material (H90Au) is limited by the film thickness, and increases with electron energy up to the value in pure Au (which is indicated at the thickness of Au film  $d_{Au} = 1200$  nm). The higher values of H90Au at the lowest thickness of Au film are given by the cell size value, because automatic division of interaction volume cannot position a cell exactly on the film/substrate

interface. The depth of the interaction volume in the Si substrate ( $H90Si$ ) decreases from the value in pure Si due to the scattering in the Au film, which increases with Au film thickness. It is clear that at some thickness of the film no excitations proceed in the substrate; in this case  $H90Si$  reaches the zero value (though in the non-zero case the film thickness is added to  $H90Si$  - see the definition in Figure 13).

The diameter of the interaction volume in the Au film ( $D90Au$ ) (Figure 15b) starts to increase relatively quickly with film thickness, and at a certain value of film thickness (depending on the electron energy) the rate of growth is suppressed and the  $D90Au$  remains constant (and approximately equal to the bulk Au value). The diameter of the interaction volume in substrate  $D90Si$  also increases with film thickness in the observed thickness range, due to the stronger elastic scattering of electrons in Au than the scattering in Si. The sudden fall to the final zero value appears at a thickness equal to the maximum depth of penetration of electrons at a given primary electron energy (see the case at electron energy 5 and 10 keV). At this thickness the Si X-ray produced intensity also falls to zero. The dimensions of the interaction volume in the substrate do not reflect in absolute scale the change in the excited X-ray intensity, which decreases substantially as the film thickness grows.

#### 4. Discussion

Elastic peak electron spectroscopy (EPES) provides an opportunity both to study electron-matter interactions under special conditions of low and medium electron energy and also to define the material parameters near the surface. In most of works on this method, the relative EPES is utilized, where firstly the experimental values of elastic reflection coefficient is compared with Monte-Carlo calculated values for one element, at one electron energy and several supposed values of IMFP. The proper value of IMFP as then found by interpolation and IMFPs for other elements can be found by measurement of electron elastic reflection. By this way, a comparison of the obtained values of IMFP with theoretical values in general can give only a measure of the influence of the experimental conditions on the IMFP values.

We suggested a way to evaluate electron elastic reflection experiments, which is in principle similar to absolute EPES. The absolute EPES could give for any element the „experimental“ value of Bethe stopping power  $S_B$  and IMFP. Our method use experiment to find the optimal value of fitting parameter  $K$  using MC calculated and experimental  $R_e(E)$  in some experimental energy range and calculation of  $R_e(E)$  dependence by MC using these fitted values of  $K$ ; the resulting value of  $K$  define the IMFP. Nevertheless, our calculations give the “integral” values of IMFP, where the *a priori* energy dependence of these quantities is supposed.

Conversely, the IMFP values obtained taking into account the surface energy losses (SEL) are substantially higher than both the NEL case and the theoretical values from TPP-2 (Fig. 6). In view of the presence of surface energy losses in the experiment, the relatively good agreement of theoretical and experimental+MC results without taking into account the surface energy losses seems not to be satisfactory. Formerly, it was assumed that in this case the resulting IMFP should be lower than the IMFP values obtained using the TPP-2 formula (Werner, 2001b). After switching to the next mechanism of energy losses, the elastic electron reflection decreases at the same IMFP. This is equivalent to the situation when, if we want to hold the experimental  $R_e$  values, IMFP must increase. From the theoretical presumptions in (Chen, 1996), (Vičánek, 1999), (Werner et al., 2003) it is known that the surface effect appears

near to the surface and the volume interactions are damped. This causes an increase of near surface IMFPs, given by volume interactions. The decrease in the volume interactions is partially compensated by the increase in the surface interactions. The provisional theory supposes that the increase in surface interactions compensates the decrease in volume interactions, thus the overall common effect is similar, as the volume interactions do not change at the surface; even though some surface interactions start after electrons have left the solid.

Our results can be explained by this model, if we adopt some more precise definitions of the obtained IMFPs. A simulation without taking the surface energy losses into account (NEL) shows the whole response of the solid, without separating the various sources of energy losses. According to theoretical suppositions, this response should be approximately equal to the bulk situation, which is proved by the good agreement of our values of IMFP, calculated using NEL condition, with TPP-2 values. In this way, we can take these values of IMFP as some „effective“ values in the vicinity of the surface. The calculated increased IMFPs, obtained under SEL conditions, show, in our opinion, the „real“ value of IMFP very near to the surface. Because in this case the surface interactions are taken into account separately, we suppose that the increase in IMFPs in comparison with TPP-2 can be explained by a real decrease in volume interactions near the surface, where most of the interactions proceed. Nevertheless, in this case the situation is rather complicated by the fact that the electron can move in various depths individually for each electron. The size of the interaction volume in which most of the scattering proceeds at low electron energy, can be obtained using MC calculation. Except of mean depth, the information depth, i.e., the depth from which 90% of elastically reflected electrons are coming from and the depth at which their last collision on an individual path took place, is calculated by our MC code. The mean depths of elastic collisions, the information depths and relative number of electrons, for which the deepest collision was in depth equal or less than 0.5 nm, for Al and Au and for several energies can be found. From this results we can see that 90% of the elastically reflected electrons come from the surface layer 3.04 and 4.2 nm in thickness (i.e., this is the information depth), respectively, for Al and 1.41 and 1.54 nm, respectively, for Au, for NEL and SEL, respectively, at 3 keV. Moreover, it is also shown that the relative number of backscattered electrons from the layer between the surface and a depth of 0.5 nm is relatively high both for Al and for Au, at least at low energy. As a consequence, the influence of surface vicinity on IMFP should decrease as the energy increases, when the depth of penetration increases. Our fitting method takes into account the whole energy region between 0.2 and 3 keV, without distinguishing this energy dependence.

There are several other possible reasons of differences between our values of IMFP and values from references. Due to the use of simple corrections of  $S_B$ , the differences of IMFP found by our method and the theoretical values from, mainly in the low energy region, could be caused by a difference between the corrected  $S_B$  and the theoretical values obtained using various more sophisticated stopping power calculations. If we compare the  $S_B$  values obtained by Bethe formula corrected for low energy with theoretical values, there are few differences between RSW and JL corrections of the Bethe stopping power for light elements (e.g., Al). In general, the JL corrected values of  $S_B$  are very near to the theoretical values given in (Ashley, 1988), (Ding & Shimizu, 1989) and (Fernandez-Varea et al., 1993). For Cu, our best fit gives the best agreement for values of  $S_B$  from (Fernandez-Varea et al., 1993). For heavier elements (e.g., Au), there are remarkable differences among various authors, and our stopping power calculated using the RSW correction (and giving the best agreement of

the MC result with experimental values of  $R_e(E)$  agrees best with that calculated in (Fernandez-Varea et al., 1993). Nevertheless, the corrected  $S_B$  is in reasonable overall agreement with theoretical suppositions of stopping power for the observed materials.

Next, the experimental values of  $R_e$  at energies  $E > \sim 1$  keV in (Schmid et al., 1983) are usually higher than the calculated values of  $R_e$  for all the elements. We suppose the reason could lie in the measurements using the RFA analyser. The relative resolution of the analyser used in this experiment increases with energy up to 5 eV at 2.5 keV (Schmid et al., 1983); in subsequent experiments (Dolinski et al, 1988a, 1988b and 1992), the resolution is estimated to be 0.6 %, i.e., 12 eV at 2 keV. Thus, some inelastically scattered electrons with low energy loss are treated as elastically scattered electrons. The surface plasmon of 3.4 eV is excited in Ag, which presence also could explain the large RMS for this element. In the calculation, all these electrons are excluded from the obtained value of  $R_e$ . By this way, the absolute EPES could give the "experimental" values of IMFP, but the measured values for usual measurements by RFA analysers are very strongly influenced by decreasing analyser resolution for higher energies.

The next disadvantage of EPES method for definition of IMFP is the influence of IMFP value on possibility or intensity of surface excitations. As a consequence, the influence of surface vicinity on IMFP should decrease as the energy increases, when the depth of penetration increases. Our fitting method takes into account the whole energy region between 0.2 and 3 keV, without distinguishing this energy dependence.

## 5. Conclusion

In simulation of EPES experiment, we suggested a way to evaluate electron reflection experiments, which is in principle similar to absolute EPES. The particular results are as follows:

- In the model, there is only one fitting parameter  $K$ , which must be determined by comparison with measurement.
- Reasonably good agreement of calculated and experimental values of the elastic electron reflection coefficient was obtained in the energy region 0.2 -  $\sim 3$  keV for the observed materials and suitable models.
- Selection of the best correction of the stopping power depends on the material.
- Incorporating surface energy excitations for MC calculations improves the agreement of  $R_e$  remarkably.
- The TFD model of elastic scattering is proved to be preferable for Al, Cu, Ag and Au.
- For IMFP, relatively good agreement with data in the literature was obtained without taking the surface excitations into account; in this case we obtained the "effective" IMFP near the surface, which defines the full response of a solid. Dividing the volume and surface excitations, we suppose that we obtained the real but "averaged" IMFP in the interaction volume near the surface, which is higher than the "effective" IMFP due to the decreasing influence of volume excitations. The depth dependence of this "averaged" IMFP cannot be determined at our way of evaluation.
- Using IMFP values from the literature for MC simulation of  $R_e$  dependencies, i.e., doing *ab initio* calculations, a comparison with experimental values indicates the medium or low influence of surface excitations, mainly for heavier materials. This proves that our method of IMFP's finding from experiment is relatively more efficient.

- Finally, the good agreement of the MC calculations with experimental values encourages the application of this fitting method to the study of IMFP and, generally, to the study of electron-matter interactions near the surface. For example, these calculations could be used to obtain the depth distribution of electron elastic collisions in solids.

Thus, our model of inelastic scattering might be reasonably appropriate for MC simulation of electron scattering by matter and for evaluation of IMFP from elastic electron reflection experiment at low electron energies.

Also the relatively good agreement in simulation of electron transmission through the thin films at the start of our work was obtained (the best using Hartree-Fockova atomic model for calculation of differential cross-sections of elastic scattering and simple hyperbolic model of electron inelastic scattering, without electron declination during scattering). Nevertheless, these simulations were realized at the start of our study, when better models were not available. In addition, with a low energy transmission electron microscope in a study of carbon, very good agreement with theoretical assumptions was found for the calculated angular distribution of elastically transmitted electrons in the given energy and thickness range. Using these results, the chromatic error was found, and in this way also the deterioration of resolution due to increasing sample thickness. Testing by measurement is necessary, but we believe the agreement of the experimental and calculated thickness and aperture dependences of transmitted electrons can be improved by implementing other models of energy losses into the code.

In our work on calculation of interaction volume, the MC data of electron backscattering and also of the k-ratios for thin films on substrate, in dependence on the energy of the electrons and on the film thickness, were compared with the experimental values, with reasonably good agreement. On the base of these findings, we have obtained reasonable estimations of the radial and vertical dimensions of the interaction volume and their dependence on the electron energy and on the film thickness. The results are important for Electron Probe Microanalysis of nonhomogeneous samples, e.g., for the analysis of films on substrates.

## 6. Acknowledgments

For a critical reading of this manuscript, for kind provision of experimental data and for helpful discussions, we would like to thank Dr. J. Zemek and Dr. J. Pavluch. This work was supported by Czech Science Foundation (GA-CR) projects No. 202/02/0237, No.101/06/0226, No. P108/10/1858 and project No. KAN101120701 of the Grant Agency of the Academy of Science of the Czech Republic. We are grateful to Mr Robin Healey for the large number of English language reviews.

## 7. References

- Ashley J C. (1988). Interaction of low-energy electrons with condensed matter - stopping powers and inelastic mean free paths from optical-data. *J. Electr. Spectr. Rel. Phenom.*, Vol.46, (1988), 199, ISSN 0368-2048
- Baró, J.; Sempau, J.; Fernandez-Varea, J.M.; Salvat F. (1995). PENELOPE - an algorithm for Monte-Carlo simulation of the penetration and energy-loss of electrons and positrons in matter. *Nucl. Instr. Meth.*, B Vol.100, (1995), p.31, ISSN: 0168-583X

- Berger, M.J. (1963). In: *Methods Computational Physics*, Adler B., (Ed.), Vol. 1, 135, Acad. Press, ISBN 0124608019, New York
- Berger, M.J.; Seltzer, J. (1964). Studies in penetration of charged particles in matter. Nucl. Sci. Ser. Rep. No. 39, NAS-NRC Publ. No. 1133, National Academy of Science, Washington DC, p. 205
- Binder, K.(1979). *Monte-Carlo Methods in Statistical Physics*, Binder K., (Ed.), Springer, ISBN 0-387-16514-2, 3-540-16514-2, Berlin , (2. edition 1986).
- Bote, D.; Salvat, F.; Jablonski, A.; Powell, C. J. (2009). The effect of inelastic absorption on the elastic scattering of electrons and positrons in amorphous solids. *Journal of electron spectroscopy and related phenomena*, Vol.175, (2009), pp.41-54, ISSN 0368-2048
- Burhop B.H.S. (1955). *Journal de physique et le Radium*, Vol.16, (1955), p.625, ISBN ISSN 1160-8161
- Cosslett, V.E.; Thomas, R.N. (1964a). Multiple scattering of 5-30 keV electrons in evaporated metals films. 1. Total transmission + angular distribution. *Br. J. Appl. Phys.*, Vol.15, (1964), p.883, ISSN 0022-3727
- Cosslett, V.E.; Thomas, R.N. (1964b). Multiple scattering of 5-30 keV electrons in evaporated metal films. 2. Range-energy relations. *Br. J. Appl. Phys.*, Vol.15, (1964), p.1283, ISSN 0022-3727
- Cosslett, V.E.; Thomas, R.N. (1965). Multiple scattering of 5-30 keV electrons in evaporated metal films. 3. Backscattering and absorption. *Br. J. Appl. Phys.*, Vol.16, (1965), p.779, ISSN 0022-3727
- Dapor, M. (2003). *Electron-Beam Interactions with Solids – Application of the Monte Carlo Method to Electron Scattering*, Springer, ISBN 3-540-00652-4, Berlin
- Delong, A.; Kolarik, V.; Martin, D.C. (1998). Low voltage transmission electron microscope LVEM-5. *14th International Congress on Electron Microscopy, Electron Microscopy 1998* Vol.1, pp.463-464, ISBN: 0-7503-0564-9, Cancun, Mexico, August 31-September 04, 1998
- Ding, Z. J.; Shimizu, R. (1989). Inelastic-collisions of kV electrons in solids. *Surf. Sci.* Vol.222, (1989), p.313, ISSN 0039-6028
- Dolinski, W.; Nowicki, H.; Mroz, S. (1988a). Application of elastic peak electron-spectroscopy (EPES) to determine inelastic mean free paths (IMFP) of electrons in copper and silver. *Surf. Interf. Anal.*, Vol.11, (1988), p.229, ISSN 0142-2421
- Dolinski, W.; Mroz, S.; Zagorski, M. (1988b). Determination of the inelastic mean free-path of electrons in silver and copper by measurement and calculation of the elastic-scattering coefficient. *Surf. Sci.*, Vol.200, (1988), p.361, ISSN 0039-6028
- Dolinski, W.; Mroz, S.; Palczynski, J.; Gruzza, B.; Bondot, P.; Porte, A. (1992). Determination of inelastic mean free-path of electrons in noble-metals. *Acta Phys Pol. A*, Vol.81, (1992), p.193, ISSN 0587-4246
- Draper, N. R.; Smith H. (1966). *Applied Regression Analysis*, Willey, ISBN 0-471-22170-8, New York.
- Fernandez-Varea, J.M.; Mayol R.; Liljequist, D.; Salvat F. (1993). Inelastic-scattering of electrons in solids from a generalized oscillator strength model using optical and photoelectric data. *J.Phys.: Condens. Matter*, Vol.5, (1993), 3593, ISSN 0953-8984
- Fernandez-Varea, J.M.; Liljequist, D.; Csillag, S.; Rätty, R.; Salvat, F. (1996). Monte Carlo simulation of 0.1-100 keV electron and positron transport in solids using optical data and partial wave methods. *Nucl. Instr. Meth. B*, Vol.108, (1996), p.35, ISSN 0168-583X

- Fernandez-Varea, J.M.; Salvat, F.; Dingfelder, M.; Liljequist D. (2005). A relativistic optical-data model for inelastic scattering of electrons and positrons in condensed matter. *Nucl. Instr. Meth. B*, Vol.229, (2005), pp.187–218, ISSN 0168-583X
- Fitting, H-J. (1974). Transmission, energy-distribution, and se excitation of fast electrons in thin solid films. *Phys. Stat. Sol. A*, Vol.26 (1974) p.525, ISSN 0031-8965
- Gauvin, R., l'Esperance, G. (1992). A Monte-Carlo code to simulate the effect of fast secondary electrons on k(ab) factors and spatial-resolution in the TEM. *Journal of microscopy-Oxford*, Vol.168, Part.2, (1992), p.153-167, ISSN 0022-2720
- Goldstein J.I., Newbury D.E., Echlin P., Joy D.C., Fiori C., Lifshin E. (1981). *Scanning Electron Microscopy and X-Ray Microanalysis: a text for biologists, materials scientists, and geologists*, Plenum Press, ISBN 0-306-40768-X, New York, p. 620.
- Chapman, J. N.; Gray, C. C.; Robertson B. W.; Nicholson W.A. P. (1983). X-ray-production in thin-films by electrons with energies between 40 and 100 keV. *X-ray Spectrometry*, Vol.12, (1983), p.153, ISSN 0049-8246
- Chen, Y.F. (1995). Effect of surface excitations in determining the inelastic mean free-path by elastic peak electron-spectroscopy. *Journal of Vacuum Science & Technology A-Vacuum Surfaces and Films*, Vol.13, (1995), p.2665, ISSN 0734-2101
- Chen, Y. F. (1996). Quantitative analysis in X-ray photoelectron spectroscopy: Influence of surface excitations. *Surf. Sci.*, Vol.345, (1996), p.213, ISSN 0039-6028
- Chen, Y. F. (2002). Surface effects on angular distributions in X-ray photoelectron spectroscopy. *Surf. Sci.* Vol.519, (2002), p.115, ISSN 0039-6028
- Inokuti, M. (1971). Inelastic collisions of fast charged particles with atoms and molecules - Bethe theory revisited. *Rev. Mod. Phys.*, Vol.43, (1971), p.297, ISSN 0034-6861
- Jablonski, A., Salvat, F., Powell, C.J. (2003). NIST electron elastic-scattering cross-section database, version 3.1. NIST: Gaithersburg, MD
- Joy, D. C.; Luo, S. (1989). An empirical stopping power relationship for low-energy electrons. *Scanning*, Vol.11, (1989), p.176, ISSN 0161-0457
- Joy, D.C. (1995). *Monte Carlo Modeling for Electron Microscopy and Microanalysis*, Oxford University Press, ISBN 0-19-508874-3, Oxford
- Jung, R.; Lee, J. C.; Orosz, G. T.; Sulyok, A.; Zsolt, G.; Menyhard, M. (2003). Determination of effective electron inelastic mean free paths in SiO<sub>2</sub> and Si<sub>3</sub>N<sub>4</sub> using a Si reference. *Surf. Sci.*, Vol.543, (2003), p.153, ISSN 0039-6028
- Kirkpatrick, P.; Wiedmann, L. (1945). Theoretical continuous X-ray energy and polarization. *Phys. Rev.*, Vol.67, (1945), p.321, ISSN 0031-899X
- Kissel, L.; Quarles, C. A.; Pratt. R. H. (1983). Shape functions for atomic-field bremsstrahlung from electrons of kinetic energy 1-500 keV on selected neutral atoms  $1 \leq Z \leq 92$ . *Atomic Data and Nuclear Data Tables*, Vol.28, (1983), p.381, ISSN 0092-640X
- Kyser D.F. (1989) in: *Introduction to Analytical Electron Microscopy*, Hren J.J., Goldstein J.I., Joy D.C., (Eds.), p.199, Plenum Press, ISBN 0-306-40280-7, New York
- Landau, L. D.; Lifshic, E. M. (1974). *Quantum Mechanics*, (translated from the Russian by J.B. Sykes and J.S. Bell). Pergamon Press, ISBN 0-08-017801-4, Oxford, p. 137
- Lednicky, F.; Coufalova, E.; Hromadkova, J.; Delong, A.; Kolarik, V. (2000). Low-voltage TEM imaging of polymer blends. *Polymer*, Vol.41, (2000), p. 4909-4914, ISSN 0032-3861
- Liljequist, D., (1978). Simplified models for monte-carlo simulation of energy-distributions of kev electrons transmitted or backscattered in various solids. *J. Phys. D: Appl. Phys.*, Vol.11, (1978), p.839, ISSN 0022-3727

- Liljequist D. (1983). A simple calculation of inelastic mean free-path and stopping power for 50 eV-50 keV electrons in solids. *J. Phys D: Appl. Phys.*, Vol.16, (1983), 1567, ISSN: 0022-3727
- Ly, T.D.; Howitt, D.G. (1992). A Monte-Carlo calculation of the backscattering coefficient for a multilayer. *Scanning*, Vol.14, (1992), p.11, ISSN 0161-0457
- Mayol, R.; Salvat, S. (1997). Total and transport cross sections for elastic scattering of electrons by atoms. *Atomic Data & Nuclear Data Tables*, Vol.65, (1997), p.55, ISSN 0092-640X
- Murata, K.; Kawata, H.; Nagami K. (1987). Electron-scattering in low-voltage scanning electron-microscope targets. *Scanning Micr. Suppl.*, Vol.1, (1987), p.83, ISSN 0891-7035
- Niedrig, H. (1982). Electron backscattering from thin-films. *J. Appl. Phys.*, Vol.53, (1982), p.R15, ISSN 0021-8979
- Oswald, R. (1992). Numerische Untersuchung der elastischen Streuung von Elektronen an Atomen und ihre Rückstreuung an Oberflächen amorpher Substanzen im Energiebereich unter 2000 eV. PhD Thesis, Tuebingen. 173 p.
- Penn, D.R. (1987). Electron mean-free-path calculations using a model dielectric function. *Phys.Rev. B* Vol.35. (1987). p.482, ISSN 0163-1829
- Pines, D. (1964). *Elementary excitations in solids: lectures on phonons, electrons, and plasmons*. Benjamin, New York, ch. 3,4. ISBN
- Powell, C.J. (1989). Cross-sections for inelastic electron scattering in solids. *Ultramicroscopy*, Vol.28, (1989), p.24, ISSN 0304-3991
- Raether, H. (1980). *Excitation of Plasmons and Interband Transitions by Electrons*. Springer Tracts in Modern Physics 88 Springer, ISBN 3-540-09677-9 - 0-387-09677-9, Berlin
- Rao-Sahib, T. S.; Wittry, D. B. (1974). X-ray continuum from thick elemental targets for 10-50 keV electrons. *J. Appl. Phys.*, Vol.45, (1974), p.5060, ISSN 0021-8979
- Reichelt, R.; Engel, A. (1984). Monte-Carlo calculations of elastic and inelastic electron-scattering in biological and plastic materials. *Ultramicroscopy*, Vol.13, (1984), p.279, ISSN 0304-3991
- Reimer, L.; Brockmann, K., Rhein, U. (1978). Energy losses of 20 – 40 keV electrons in 150 – 650  $\mu\text{g}/\text{cm}^2$  metal films. *J. Phys D: Appl. Phys.*, Vol.11, (1978), p.165, ISSN:0022-3727
- Reimer, L. (1984). *Transmission Electron Microscopy: physics of image formation and microanalysis*, Springer, ISBN 0-387-11794-6 , 3-540-11794-6, Berlin, p.138
- Reimer, L.; Lodding, B. (1984). Calculation and tabulation of Mott cross-sections for large-angle electron-scattering. *Scanning*, Vol.6, (1984), p.128, ISSN 0161-0457
- Reimer, L.; Stelter, D. (1986). Fortran-77 Monte-Carlo program for minicomputers using Mott cross-sections. *Scanning*, Vol.8, (1986) pp.265-277, ISSN 0161-0457
- Reimer, L. (1996). Monte Carlo simulation techniques for quantitative X-ray microanalysis. *Microchimica Acta [Suppl.]*, Vol.13, (1996), p.1, ISSN 0026-3672
- Riley M.E. et al. (1975), *Atomic & Nuclear Data Tables*, Vol.15, (1975), p.443, ISSN 0092-640X
- Ritchie, R.H. (1982). Energy-losses by swift charged-particles in the bulk and at the surface of condensed matter. *Nucl. Instr. and Meth.*, Vol.198, (1982), p.81, ISSN 0029-554X
- Salvat, F.; Mayol, R. (1993). Elastic-scattering of electrons and positrons by atoms - Schrodinger and Dirac partial-wave analysis. *Comput. Phys. Commun.*, Vol.74, (1993), p.358, ISSN 0010-4655

- Salvat, F.; Jablonski, A.; Powell, C. J. (2005). ELSEPA - Dirac partial-wave calculation of elastic scattering of electrons and positrons by atoms, positive ions and molecules. *Comput. Phys. Commun.*, Vol.165, (2005), p.157, ISSN 0142-2421
- Shimizu, R.; Kataoka, Y.; Matsukawa, T.; Ikuta, T.; Murata, K.; Hashimoto, H. (1975). Energy-distribution measurement of transmitted electrons and Monte-Carlo simulation for kilovolt electron. *J. Phys. D: Appl. Phys.*, Vol.8, (1975), p.820, ISSN 0022-3727
- Schmid, R.; Gaukler, H. K.; Seiler, H. (1983). Measurement of elastically reflected electrons (e-less-than-or-equal-to-2.5 keV) for imaging of surfaces in a simple ultra high-vacuum scanning electron-microscope. *Scan. Electr. Micr.*, Vol.2., (1983), 501, ISSN 0891-7035
- Schwaab, P. (1987). Quantitative energy dispersive-x-ray microanalysis of thin metal specimens using the STEM. *Scanning*, Vol.9, (1987), p.1, ISSN 0161-0457
- Starý, V. (1996). The check of the elastic scattering model in Monte-Carlo simulation. *Microscopica Acta A Suppl.*13, (1996), 559, ISSN 0026-3672
- Starý, V. (1999). Fitting a Simple Model of Inelastic Scattering in Monte Carlo Code to Experimental Data. *J. Phys. D: Appl. Phys.*, Vol.32, (1999), pp.1811-1818, ISSN 0022-3727
- Starý, V. (2000). Comparison of Monte Carlo simulation and measurement of electron reflection from solids. Proc. of MC2000 - Advanced Monte Carlo for Radiation Physics, Particle Transport Simulation and Applications, Kling A. et al. , (Eds.), p. 369, ISBN 3-540-41795-8, Lisbon, October 23-26, 2000
- Starý, V.; Jurek, K. (2002). X-ray emission from thin films on a substrate - Calculation and experiments. *Microchimica Acta*, Vol.139, (2002), p.179, ISSN 0026-3672
- Starý, V. (2003). Monte-Carlo Simulation of Electron Interaction with a Thin Film. *Thin Solid Films*, Vol.433, (2003), pp.326-331, ISSN 0040-6090
- Starý, V.; Pavluch, J.; Zemek J. (2004). Monte-Carlo Simulation of Low-energy Electron Elastic Reflection by Solids - Assessment by Experiment, *42nd IUVSTA Workshop 'Electron Scattering in Solids: From Fundamental Concepts to Practical Applications'*, abstract in: *Surface and interface analysis*, Vol.38, (2006), p.88-117, Wiley InterScience, eISSN 1096-9918; pISSN 0142-2421, Debrecen, Hungary, July 4-8, 2004
- Starý, V.; Zemek, J.; Pavluch J. (2007). Angular and energy distribution of backscattered electrons simulated by Monte-Carlo - Assessment by experiment I. *Vacuum*, Vol.82, (2007), pp.121-124, ISSN 0042-207X
- Tanuma, S.; Powell, C. J.; Penn, D. R. (1987). Proposed formula for electron inelastic mean free paths based on calculations for 31 materials. *Surface Science*, Vol.192, (1987), p.L849-L857, ISSN 0039-6028
- Tanuma, S.; Powell, C. J., Penn, D. R. (1991a). Calculations of electron inelastic mean free paths. 2. Data for 27 elements over the 50-2000 eV range. *Surface and Interface Analysis*, Vol.17, (1991), p.911, ISSN 0142-2421
- Tanuma, S.; Powell, C. J., Penn, D. R. (1991b). Calculations of electron inelastic mean free paths. 3. Data for 15 inorganic-compounds over the 50-2000 eV range. *Surface and Interface Analysis*, Vol.17, (1991), p.927, ISSN 0142-2421
- Tanuma, S.; Powell, C. J., Penn, D. R. (1993). Calculations of electron inelastic mean free paths. 4. Evaluation of calculated IMFPs and of the predictive IMFP formula TPP-2 for electron energies between 50 and 2000 eV. *Surface and Interface Analysis*, Vol.20, (1993), p.77, ISSN 0142-2421

- Tanuma, S.; Powell, C. J., Penn, D. R. (1994). Calculations of electron inelastic mean free paths. 5. Data for 14 organic-compounds over the 50-2000 ev range. *Surface and Interface Analysis*, Vol.21, (1994), p.165, ISSN 0142-2421
- Tougaard, S.(1997). Universality classes of inelastic electron scattering cross-sections. *J. Surf. Interface Anal.*, Vol.25, (1997), p.137, ISSN 0142-2421
- Tung, C.J.; Ashley, J.C.; Ritchie, R.H. (1979). Electron inelastic mean free paths and energy-losses in solids. *Surf. Sci.*, Vol.81, (1979), p.427, ISSN 0039-6028
- Vičánek, M. (1999). Electron transport processes in reflection electron energy loss spectroscopy (REELS) and X-ray photoelectron spectroscopy (XPS). *Surf. Sci.*, Vol.440, (1999), p.1, ISSN 0039-6028
- Werner, W. S. M.; Smekal, W.; Tomastik, Ch.; Stori H. (2001a). Surface excitation probability of medium energy electrons in metals and semiconductors. *Surf. Sci.*, Vol.486, (2001), p.L461, ISSN 0039-6028
- Werner, W. S. M. (2001b). Electron transport in solids for quantitative surface analysis. *Surface and Interface Analysis*, Vol.21, (2001), p.141, ISSN 0142-2421
- Werner, W. S. M.; Eisenmenger-Sittner Ch.; Zemek J.; Jiricek P. (2003). Scattering angle dependence of the surface excitation probability in reflection electron energy loss spectra. *Phys. Rev. B*, Vol.67, (2003), p.155412, ISSN 1098-0121

# Monte Carlo Simulation of SEM and SAM Images

Y.G. Li, S.F. Mao and Z.J. Ding

*Hefei National Laboratory for Physical Sciences at Microscale and Department of Physics,  
University of Science and Technology of China,  
Hefei, Anhui 230026,  
P. R. China*

## 1. Introduction

### Monte Carlo method and its applications to electron microscopic and spectroscopic techniques

#### 1.1 Electron microscopic and spectroscopic techniques

The electron microscopic and spectroscopic techniques are conveniently and widely used for surface and bulk analysis of materials. These analysis tools use the various types of electron signals that emitted from the specimen irradiated by a beam of mono-energetic primary electrons or X-rays for imaging of surface, and for structural and chemical characterization. Different signals in relevant techniques are secondary electrons (SEs) and backscattered electrons (BSEs) for scanning electron microscopy (SEM), Auger electrons (AEs) for Auger electron spectroscopy (AES) and scanning Auger microscopy (SAM), photoelectrons for X-ray photoelectron spectroscopy (XPS) and X-ray photo-emission electron microscopy (XPEEM), elastic scattered electrons for elastic peak electron spectroscopy (EPES), inelastic scattered electrons for electron energy loss spectroscopy (EELS) and reflection electron energy loss spectroscopy (REELS), characteristic X-ray and bremsstrahlung for electron probe microanalysis (EPMA) and analytical electron microscopy (AEM) (Reimer, 1998).

Particularly, SEM is more frequently used for a quick sample characterization. Imaging of microstructure of materials with SEs and BSEs plays a very important role in many scientific and technological fields. SE images, formed by SEs of very low energies (<50 eV) emitted from the surface region, provide mainly topographic information of the specimen surface with nanometer resolution with a modern SEM. BSE images can provide more information about the matrix composition for the signal electrons are transported from the sample interior within interaction volume of primary electrons of several keV energy. SAM is a technique that combines AES with SEM and is commercially available as an ultra-high vacuum instrument for chemical investigation of clean surfaces. With SAM it is possible to observe the surface elemental distribution and to obtain chemical state information by detecting AEs that carry characteristic energies representing the specific energy levels of surface atoms ionized by an incident electron beam.

The principle of these techniques relies on electron-solid interaction. Therefore, the study of electron transport is very important to these techniques for a detailed understanding of a

variety of physical processes involved in the electron–solid interaction. These physical processes comprise elastic and inelastic scattering in bulk and at surface. Such processes can occur repeatedly (multiple scattering), so that a combination of scattering effects that are responsible for some important features is observed in experimental spectra or images.

### 1.2 Monte Carlo method

The Monte Carlo (MC) method was initiated in the 1940s by Ulam and von Neumann who were working on Manhattan project in Los Alamos; they considered to design a novel numerical method with the use of random numbers to solve the problem of neutron transport (Metropolis, 1987). Nowadays, MC methods are widely used in many fields to solve complex physical and mathematical problems (James, 1980; Rubinstein, 1981; Kalos & Whitlock, 1986), particularly those involving multiple independent variables where other numerical methods would demand formidable amounts of memory and computing time.

MC electron trajectory simulation method has been used since 1950's to electron probe microanalysis, electron spectroscopy and electron microscopy, for obtaining quantitative information on different signals recorded by these instruments. Various MC physical models have been proposed and used for specific purposes. The main advantages of MC simulation are the easy implementation of different scattering channels with their cross sections into a simulation model and the ability to describe radiation transport through material structures with complex geometry boundaries.

In a MC simulation of electron transport, an electron trajectory is tracked as a random sequence of flights that end with a scattering event where the electron changes its direction of movement and/or loses energy, and produces secondary signals in an inelastic event. For a given experimental condition numerical random trajectory histories of electrons have to be simulated to present statistically meaningful calculation result. To such a simulation a MC physical model is essential, which considers how to treat electron scattering with corresponding formulation, e.g. a set of differential cross sections (DCS), for the relevant interaction mechanism. The DCSs determine the probability distribution functions (PDF) of the random variables that represent physical quantities for tracing an electron track, e.g. free path between successive events, type of interaction taking place, energy loss and/or angular deflection in a particular scattering event. Once these PDFs are known, random sampling can be made so that a trajectory history is formed. When a large number of trajectory histories are generated, quantitative information on the signal transport process may be obtained by a statistical averaging over the simulated histories (Salvat et al., 2006).

A MC simulation yields the similar information as the solution of Boltzmann transport equation, with the same interaction model, but is easier to be implemented (Berger, 1963). In particular, the MC simulation of radiation transport in a sample with complex geometry is straightforward, while even the simplest finite geometries are very difficult to be dealt by transport equation method. The drawback of the MC method lies in its random nature, that is, the calculation results suffers statistical uncertainties, which can be reduced at the expense of increasing sampling population, and, hence, the computation time.

### 1.3 Physical processes

The present status of the MC calculation, particularly related to SEM/SAM, is outlined. For the treatment of electron elastic scattering, both the screened Rutherford formula and the Mott differential cross section have been available. Since it was been found (Ichimura & Shimizu,

1981; Reimer & Krefting, 1976) that the use of the Mott cross section is more satisfactory than the Rutherford cross section, particularly for heavier elements and at lower energies, the employment of the Mott cross sections in the keV and sub-keV energy region is now popular. Regarding the approach to electron inelastic scattering, the Bethe stopping power equation in the continuous slowing-down approximation (CSDA) has been widely used with considerable success. To include fast secondary electron generation, some modifications have also been made to the CSDA: the hybrid of the CSDA with the individual energy loss processes due to inner-shell ionization (Ichimura & Shimizu, 1981), the utilization of the discrete inelastic scattering cross section of Moller (Moller, 1931; Murata et al., 1981), and the use of generalized oscillator strength in a hydrogenic approximation for inner shells (Desalvo & Rosa, 1987), each of which permits the simulation of the fast knock-on electrons. However, any characteristic energy loss process specific to a sample is omitted under the CSDA. Considering that Bethe's equation is valid only at sufficiently high electron energies, Rao-Sahib & Wittry (Rao-Sahib & Wittry, 1974) have empirically extrapolated Bethe stopping powers to the low energy region by assuming a parabolic function,  $-dE/ds \propto E^{-1/2}$ , which has been extensively used in the simulation of the slowing down process of slow electrons (Joy, 1987; Kotera, 1989; Luo et al., 1987; Newbury et al., 1990). But this formulation gives energy dependence opposite of that predicated by the Lindhard theory for free electron gas (Lindhard, 1954; Ritchie et al., 1969) and overestimates significantly the energy loss of low-energy electrons. Concerning secondary electron generation, both models of the secondary electron excitation assumed from the stopping power formula (Joy, 1987; Matsukawa & Shimizu, 1974; Murata et al., 1987) or from the Streitwolf (1959) equation (Koshikawa & Shimizu, 1973; Kotera, 1989; Kotera et al., 1990; Luo et al., 1987) have required fitting parameters in order to get the correct secondary electron yield. Furthermore, the secondary emission process was simply described by an exponential decay law (Joy, 1985) or was hybridized with a cascade model of secondary production (Luo & Joy, 1990) and with emission processes (Koshikawa & Shimizu, 1973). Another unsatisfactory situation is that one usually divides the energy region for fast secondary electrons and the low secondaries to adopt different approaches for each (Ding & Shimizu, 1988a; Ding & Shimizu, 1989a; Kotera, 1989) because the available models of electron scattering and secondary generation were limited in certain energy range. Therefore, a unified treatment of electron inelastic scattering and secondary electron generation is quite necessary. Perhaps the best approach should be based on a dielectric function which characterizes the specific excitation processes of a sample (Pines, 1964). A dielectric function  $\varepsilon(\mathbf{q}, \omega)$  can provide us with detailed knowledge of energy loss cross section and scattering angular distribution for electron inelastic scattering. This has been achieved (Cailler & Ganachaud, 1990; Ganachaud & Cailler, 1979a; Ganachaud & Cailler, 1979b) for free electron metal, Al, using the well-known Lindhard dielectric function describing the plasmon excitation and electron-hole pair production. Unfortunately, the ideal Lindhard dielectric function of free electron gas in the random phase approximation (Fetter & Walecka, 1971) is valid only for limited materials, that is, so-called free electron metals, and is hardly applicable to other materials such as transition and noble metals, for which the optical dielectric data have shown complexities due to interband transitions (Rather, 1980). Some modified analytical dielectric functions in the plasmon-pole approximation with damping (Brandt & Reinheimer, 1970) were also limited to materials, for which the damping plasmon dominates the energy loss processes of electrons, such as carbon and silicon (Desalvo et al., 1984). Furthermore, the theoretical calculation of  $\mathbf{q}$ -

dependent dielectric function,  $\varepsilon(\mathbf{q}, \omega)$ , is difficult and has been numerically evaluated (Nizzoli, 1978; Singhal, 1975; Sramek & Cohen, 1972; Walter & Cohen 1972) only for selected  $\mathbf{q}$ s and the first few reciprocal lattice vectors using realistic band structure data for some simple metals and semiconductors (Sturm, 1982). In fact, the comprehensive first principle theoretical calculations for metals have been limited to  $\varepsilon(\omega) = \varepsilon(\mathbf{q} = 0, \omega)$  (Maksimov et al., 1988). It is thus impractical to use  $\varepsilon(\mathbf{q}, \omega)$  derived from a band structure calculation for a MC simulation, hence we have to use the optical dielectric data which are available experimentally from optical method and electron energy loss spectroscopy (Egerton, 1986). Systematic data of the dielectric constants have been provided and compiled for a number of materials for practical use (Hagemann et al., 1975; Palik, 1985; Palik, 1991) with advance in use of synchrotron radiation facilities.

Our first attempt (Ding & Shimizu, 1988a) had used the approach given by Powell (Powell, 1985). The  $\mathbf{q}$ -integrated excitation function for electron energy loss and production of secondary electrons is related to  $\varepsilon(\omega)$  and a parameter, which was determined by fitting the calculated electron mean free path with experimental data. Reasonable accuracy has been achieved in the calculation of the energy distribution of backscattered electrons (Ding et al., 1988b). However, the angular information in electron inelastic scattering was accumulated by the integration over  $\mathbf{q}$ , and the parameter-involved approach is not favorable for general use. Particularly, this parameter does not allow describing the Bethe stopping powers at high energies, so that we had to use this simple dielectric model only for slow electrons.

According to Penn's work (Penn, 1987), the  $\mathbf{q}$ -dependent electron energy loss function may be derived from optical dielectric constants. This algorithm enables us to calculate energy loss cross section and scattering angular distribution required for a MC simulation of discrete electron inelastic scattering processes (Ashley, 1991; Ding & Shimizu, 1989b). It has been shown that the method yields the Bethe stopping powers at high energies (Ashley, 1988; Ding & Shimizu, 1989b), and the calculated electron mean free paths fit the experimental data in a wide energy region for many elements and compounds. This fact indicates that the dielectric function modeling is very useful for MC simulation of electron inelastic scattering. We have used it in the calculation of x-ray depth profiles (Ding & Wu, 1993) and background in Auger electron spectroscopy (Ding et al., 1994), and Tokesi et al. (Tokesi et al., 1996) have calculated the reflected electron energy loss spectrum. Jensen & Walker (Jensen & Walker, 1993) have also employed it to study the backscattering yield of positrons and electrons of high energies, but they failed to get good agreement with experimental data for electrons, partly because of the neglect of the secondary production. We shall demonstrate, by including cascade secondary electron generation that the backscattering yields as well as the angular energy distribution describe the precise experimental curve down to low energies very well. This modeling describes the inelastic scattering reasonably well, covering the wide energy range from several eV above the Fermi energy to several tens keV. Furthermore, the simulation of cascade production of secondary electrons included with discrete electron inelastic collisions can be directly made in a simple way.

Furthermore, great efforts have also been made for the structures simulation with complex geometries near term years. Some studies using a MC electron-trajectory simulation technique have been carried out for several kinds of simpler geometrical specimens (Gauvin, 1995; Hovington et al., 1997a; Hovington et al., 1997b; Drouin et al., 1997; Ly et al., 1995; Radzinski & Russ, 1995; Howell, 1996; Lowney, 1995; Lowney, 1996; Postek et al., 2002; Seeger et al., 2003; Yan & El Gomati, 1998). Gauvin (Gauvin, 1995) performed

simulations of X-ray images and BSE images for a spherical inclusion of homogeneous composition embedded in a matrix, based on use of Mott elastic-scattering cross sections and a modified continuous slowing-down approximation. Their CASINO program is a single-scattering MC simulation of electron trajectories in a solid, specially designed for the interactions of low-energy electrons in bulk solids and thin foils, and can be used to generate the usual recorded signals in a SEM (X-rays, SEs, and BSEs) either for point analysis, line-scans, or images (Hovington et al., 1997a; Hovington et al., 1997b; Drouin et al., 1997). Ly et al. simulated SEM images of spheres of different materials on a substrate surface and at various depths beneath the flat surface (Ly et al., 1995). Radzimski and Russ performed simulations of BSE images of three-dimensional (3D) multilayer and multi-element structures, on the basis of a single-scattering procedure, to study electron beam and detector characteristics (Radzimski & Russ, 1995). Their simulation procedure also took into account the effects of the electrical and angular characteristics of a solid-state detector and the effect of the electron beam size on image quality and certain artifacts. Howell et al. developed a program to illustrate macro topographies on electron backscattering (Howell, 1996). This program can simulate a target constructed with a choice of a flat surface, a circular filament, or a rough surface simulated by a sine wave. Another MC simulation program, MONSEL, has been used to model the interaction of an electron beam with one or two lines lithographically produced on a multilayer substrate (Lowney, 1995; Lowney, 1996). The simulated signals include transmitted, backscattered and secondary electrons. Another application of this program was concerned with the simulation of two-dimensional (2D) SE and BSE images of a simple notch (Postek et al., 2002). Recently, the MONSEL program has been extended by Seeger et al. to simulate SE and BSE images of a complex structure consisting of many triangles to create a complex specimen surface (Seeger et al., 2003). Certain programming techniques enabled faster calculations. Yan and Gomati developed a 3D MC code to simulate images of BSEs and AEs for a more complex specimen (Yan & El Gomati, 1998). Their code required the 3D geometric structure to be described in analytic form.

Because of the difficulty of simulating secondary electron generation and emission processes for specimens with complex structures, most previous studies emphasized simulation of BSE images with very simple structures, and only a few were capable of simulating SE images for specimens with complex structures. Also, inhomogeneous distribution of chemical composition inside a sample has not been considered generally. For this, the constructive solid geometry (CSG) modeling (Ding & Li, 2005; Li & Ding, 2005; Yue et al., 2005) and finite element triangle mesh modeling (Ding & Wang, unpublished; Li et al., 2008) have been developed to construct arbitrary geometric structure. Also, the inhomogeneous distribution of chemical composition inside a complex structure has been used in the subsequent researches (Yue et al., 2005; Li et al., 2009).

In this respect, the MC model (Li et al., 2008) used here for SEM/SAM image simulation has been improved in three aspects with respect to our previous simulation models (Ding & Shimizu, 1996; Ding & Li, 2005; Li & Ding, 2005; Shimizu & Ding, 1992; Yue et al., 2005): First, the full Penn dielectric function (Mao et al., 2008) is employed for the treatment of electron inelastic scattering to replace single-pole approximated (SPA) dielectric function. Second, we combine the constructive solid geometry modeling (Ding & Li, 2005; Li & Ding, 2005; Yue et al., 2005) and finite element triangle mesh modeling (Ding & Wang, unpublished; Li et al., 2008) to construct an arbitrary geometric structure. Third, we use a ray-tracing technique (Ding & Li, 2005; Li & Ding, 2005; Yue et al., 2005) for an

inhomogeneous specimen with a complex geometric structure and introduce the space subdivision method to accelerate the calculation (Ding & Wang, unpublished; Li et al., 2008). Additionally, a rough surface geometry model is introduced to construct the sample surface, together with using of a ray-tracing technique (Li & Ding, 2005) in the calculation procedure of electron step length. Appropriate boundary correction (Yue et al., 2005) had also been considered in order to treat the reflection/refraction of low energy secondary electrons when they pass through an interface separating different materials. The present MC simulation model, therefore, is probably most useful for application to SEM/SAM. In Sec. 2, detailed physical/mathematical model of electron transportation, MC simulation method and complex construction algorithms will be introduced. And then the applications for simulation of CD-SEM images for critical dimension (CD) nanometrology and simulation study of SAM images will be demonstrated in Sec. 3 and 4, respectively.

## 2. Monte Carlo modeling

Programming a simulation code for a study of electron-solid interaction relies on knowledge about two aspects: first, the theoretical description of electron scattering in the solid and cascade process; second, the reasonable description of the sample geometry boundary with correction to the MC procedure. A basic MC model for simulation on a semi-infinite sample with flat surface can be established based on well understanding of the first aspect; the MC procedure deal with the essential processes of the interaction between incident electrons and solid. To implement the MC simulation for a sample with complex geometry, the modeling of complex sample geometry becomes indispensable.

### 2.1 Physical modeling of electron transport

The interaction between the electron and solid consists of three elemental physical processes: elastic scattering, inelastic scattering and cascade electron generation. The electron elastic and inelastic scattering dominate the deflection and energy loss of electrons, respectively. Following electron inelastic scattering events, the energy transferred from an electron to a solid may induce the generation of cascading electrons (excitation of solid electrons) and this process will lead to the generation of signal electrons, especially SE signals. In a MC model for a semi-infinite sample with flat surface, the reasonable description of these three elemental processes is demanded. A detailed theoretical algorithm to deal with interaction processes is described below and a description of the MC procedure is also given.

#### 2.1.1 Elastic scattering

When a moving electron meets the positively-charged nucleus, it may be deflected without energy loss (ignored the recoil energy) due to the electric interaction. This phenomenon is called elastic scattering, which could be described by elastic scattering cross section in unit of area. Rutherford equation is the classical formulation of differential elastic scattering cross section with scattering angle,

$$\frac{d\sigma_e}{d\Omega} = \frac{Z^2 e^4}{4E^2 (1 - \cos\theta + 2\beta)^2}, \quad (1)$$

where  $Z$ ,  $e$ ,  $E$  and  $\theta$  are the atomic number, electron charge, kinetic energy of the moving electron and scattering angle, respectively.  $\beta$  is a screening parameter used to include the

influence of the atomic electron cloud. Rutherford elastic scattering cross section has been widely used in the MC simulation in past year because of its simplicity. However, the Rutherford elastic cross section shows its limitation when it applies for slow electron and/or heavy element (Walker, 1971). Furthermore, considering the wave properties of electrons, more precise elastic scattering cross section should be derived in the quantum picture instead of classical picture (Ding & Shimizu, 2003).

According to Mott (Mott, 1929), by solving Dirac equation a relativistic representation of the differential elastic scattering cross section is given by,

$$\frac{d\sigma_e}{d\Omega} = |f(\theta)|^2 + |g(\theta)|^2, \quad (2)$$

where the scattering amplitudes is derived by a partial wave expansion method (Mott & Massey, 1965):

$$\begin{aligned} f(\theta) &= \frac{1}{2ik} \sum_{l=0}^{\infty} \left[ (l+1) \left( e^{i2\delta_l^+} - 1 \right) + l \left( e^{i2\delta_l^-} - 1 \right) \right] P_l(\cos\theta), \\ g(\theta) &= \frac{1}{2ik} \sum_{l=1}^{\infty} \left( -e^{i2\delta_l^+} + e^{i2\delta_l^-} \right) P_l^1(\cos\theta), \end{aligned} \quad (3)$$

where  $\hbar k$  is the electron momentum,  $P_l(\cos\theta)$  and  $P_l^1(\cos\theta)$  are Legendre and the first order associated Legendre functions,  $\delta_l^+$  and  $\delta_l^-$  are the phase shifts of the  $l$ th partial wave for spin up and spin down electrons, respectively. A detailed numerical technique for calculation the phase shift could be found in Yamazaki's work (Yamazaki, 1977) which follows that of Bunyan & Schonfelder (Bunyan & Schonfelder, 1965). There are some published databases (Fink & Yates, 1970; Fink & Ingram, 1972; Gregory & Fink, 1974; Mayol & Salvat, 1997) in which the phase shifts and differential elastic scattering cross section have been tabulated. The most recent database is released by National Institute of Standards and Technology (NIST) (<http://www.nist.gov/srd/nist64.htm>) for providing differential and total elastic electron scattering cross sections, phase shifts and transport cross sections.

A comparison between Mott and Rutherford elastic scattering cross sections for Au at electron energy of 400 eV is shown in Fig. 1. Some obvious structures appear at larger scattering angles, which have been verified through a comparison with experimental measurements performed on Au vapor (Ding, 1990). They may be smoothed at high energies because a large number of partial waves are involved. According to a systematic comparison between these two cross sections (Ichimura, 1980; Shimizu & Ichimura, 1981), Mott elastic cross section is shown more accurate for heavy atoms and for slow electrons because spin-orbit interactions, which is dealt with more reasonably in Eq. (2), are important.

The total elastic cross section  $\sigma_e$  can be obtained by integrating the differential elastic scattering cross section over whole solid angles,

$$\sigma_e = \int \frac{d\sigma_e}{d\Omega} d\Omega = 2\pi \int_0^\pi \sin\theta \left\{ |f(\theta)|^2 + |g(\theta)|^2 \right\} d\theta, \quad (4)$$

which is related to elastic scattering mean free path in solid via

$$\lambda_e = (N\sigma_e)^{-1}, \quad (5)$$

where  $N = N_A\rho/A$  is the density of atoms,  $N_A$  is Avogadro's number,  $\rho$  is the density and  $A$  is the atomic weight. The elastic scattering mean free path is the average distance between two successive elastic collisions between a moving electron and solid atoms, which is the basic information requested in the MC simulation for electron transport in solids.

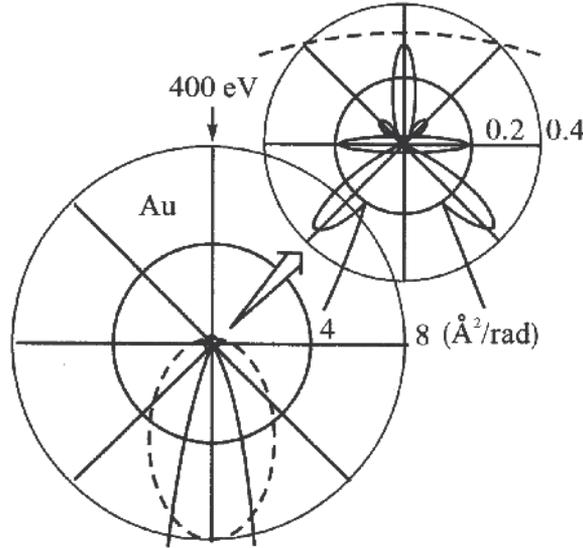


Fig. 1. Polar plot of differential electron elastic scattering cross sections for Au atom at electron energy of 400 eV. Mott and Rutherford cross sections are shown by solid and dashed lines, respectively.

### 2.1.2 Inelastic scattering

The moving electron can also encounter collisions with the solid electrons, accompanied by energy loss and excitation of solid electrons. This type of collisions involving energy transfer is called inelastic scattering, and can be described by the differential inverse inelastic mean free path (Ding, 1990),

$$\frac{d^2\lambda_{in}^{-1}}{dq d\omega} = \frac{\hbar}{\pi a_0 E} \text{Im} \left\{ \frac{-1}{\varepsilon(q, \omega)} \right\} \frac{1}{q}, \quad (6)$$

where  $\lambda_{in}$  denotes inelastic mean free path (IMFP),  $\hbar q$  and  $\hbar\omega$  are momentum transfer and energy loss, respectively;  $a_0$  is Bohr radius and  $\hbar$  is Planck constant.  $\text{Im}\{-1/\varepsilon(q, \omega)\}$  is the energy loss function and  $\varepsilon(q, \omega)$  is the dielectric function. In Eq. (6), the energy loss function is the only unknown quantity. Once the energy loss function is determined, the electron IMFP can be theoretically evaluated by the integration

$$\lambda_{in}^{-1} = \int_0^{E-E_F} d\omega \int_{q_-}^{q_+} dq \frac{d^2\lambda_{in}^{-1}}{dq d\omega}, \quad (7)$$

where the integration limits,  $\hbar q_{\pm} = \sqrt{2m}(\sqrt{E} \pm \sqrt{E - \hbar\omega})$ , are the largest and the smallest momentum transfers kinematically allowed for given  $E$  and  $\omega$ . The restriction to  $\hbar\omega \leq E - E_F$  is due to the Pauli exclusion principle that an electron can not fall into the Fermi sea which is already occupied by electrons in the solid.

It is difficult to give a theoretical calculation for the energy loss function because of its complexity. The energy loss function includes the contributions of plasmon excitation, valance electron excitation, inner-shell electron excitation and so on. However, considering the fact that the optical energy loss function  $\text{Im}\{-1/\varepsilon(\omega)\}$  is the limit of energy loss function as  $q \rightarrow 0$ , an alternative way, by extrapolating optical energy loss function to  $(q, \omega)$ -plane, can be applied to theoretically evaluate energy loss function. The advantage of this method is that abundant optical data for  $\varepsilon(\omega)$  in the loss energy range of  $10^0 - 10^4$  eV are available and are compiled in a database (Palik, 1985; Palik, 1991); use of these data derived from experiments allows a more accurate description of electronic excitation in real materials. Assuming the statistical approximation by neglecting the vertex correction, self-consistency, exchange and correlation and considering that the charge distribution in the Wigner-Seitz cell is spherically symmetric, Penn (Penn, 1987) has proposed an extrapolation method by expanding the energy loss function in terms of the Lindhard energy loss function without using any fitting parameters:

$$\text{Im}\left\{\frac{-1}{\varepsilon(q, \omega)}\right\} = \int_0^{\infty} d\omega_p g(\omega_p) \text{Im}\left\{\frac{-1}{\varepsilon_L(q, \omega; \omega_p)}\right\}, \quad (8)$$

where the expansion coefficient  $g(\omega)$  is related to the optical energy loss function by

$$g(\omega) = \frac{2}{\pi\omega} \text{Im}\left\{\frac{-1}{\varepsilon(\omega)}\right\} \quad (9)$$

and  $\varepsilon_L(q, \omega; \omega_p)$  is the Lindhard dielectric function (Lindhard, 1954) of the free electron gas with plasmon energy  $\hbar\omega_p$  in the long-wave limit,

$$\varepsilon_L^r = 1 + \frac{2}{\pi a_0 q} \frac{1}{Z} \left[ \frac{1}{2} + \frac{1}{8Z} F\left(Z - \frac{X}{4Z}\right) + \frac{1}{8Z} F\left(Z + \frac{X}{4Z}\right) \right], \quad (10)$$

$$\varepsilon_L^i = \begin{cases} \frac{1}{8a_0 k_F} \frac{X}{Z^3} & , 0 \leq X \leq 4Z(1-Z); \\ \frac{1}{8a_0 k_F} \frac{1}{Z^3} \left[ 1 - \left( Z - \frac{X}{4Z} \right)^2 \right] & , |4Z(1-Z)| \leq X \leq 4Z(1+Z); \\ 0 & , \text{otherwise,} \end{cases} \quad (11)$$

where  $F(x) = (1-x^2) \ln|(x+1)/(x-1)|$ ,  $X = \hbar\omega/E_F$ ,  $Z = q/2k_F$ .  $E_F = \hbar^2 k_F^2 / 2m$  is the Fermi energy and  $k_F$  is the Fermi wave vector. They are related to the plasmon energy through the electron density.  $\varepsilon_L^r$  and  $\varepsilon_L^i$  denote the real and the imaginary parts of the Lindhard dielectric function, respectively.

The above extrapolation method is called the full Penn algorithm (FPA) here. In the same paper Penn has further introduced the well-known single-pole approximation (SPA) to simplify the calculation. The approximation was indeed useful for the initial guide to find the trend of the energy dependence of IMFP, so that many other calculations followed the same route (Ding & Shimizu, 1988; Tanuma et al., 1988; Tanuma et al., 2005). The detail implementation of the FPA and SPA is given below, together with a comparison between them, which indicates the validity of SPA.

The implementation of FPA is carried out as follows. For a free electron gas with a certain electron density or plasmon frequency,  $\omega_p$ , only the area in the  $(q, \omega)$ -plane along the plasmon dispersion line, where  $\varepsilon_L^r = 0$  and  $\varepsilon_L^i = 0$ , can have a contribution to the integration in Eq. (8); the Lindhard energy loss function along this plasmon dispersion line is divergent; while for  $\varepsilon_L^i \neq 0$ , only the single electron excitation is allowed. Then we need to divide the calculation into two parts, i.e. single electron excitation and plasmon excitation in the area  $\varepsilon_L^i \neq 0$  and  $\varepsilon_L^i = 0$ , respectively (Jensen & Walker, 1993).

The single electron excitation part of the calculation is performed directly with Eqs. (8)-(11),

$$\text{Im} \left\{ \frac{-1}{\varepsilon(q, \omega)} \right\}_e = \int_0^\infty d\omega_p g(\omega_p) \text{Im} \left\{ \frac{-1}{\varepsilon_L(q, \omega; \omega_p)} \right\} \Theta \left[ q^+(\omega; \omega_p) - q \right] \Theta \left[ q - q^-(\omega; \omega_p) \right]. \quad (12)$$

Here  $\text{Im} \left\{ -1/\varepsilon(q, \omega) \right\}_e$  represents the energy loss function of the single electron excitation part, and  $q^-$  and  $q^+$  are the left and right boundaries of the area for  $\varepsilon_L^i \neq 0$ , respectively,

$$\begin{cases} q^-(\omega; \omega_p) = -k_F(\omega_p) + \sqrt{k_F^2(\omega_p) + 2m\omega/\hbar} \\ q^+(\omega; \omega_p) = k_F(\omega_p) + \sqrt{k_F^2(\omega_p) + 2m\omega/\hbar} \end{cases}. \quad (13)$$

Note that in Eq. (12) the plasmon frequency  $\omega_p$  is used as a variable which scans over the loss energy range of the available optical data. For the plasmon excitation part, because of the existence of the positive infinity, the integration of  $\omega_p$  can be removed to give the energy loss function of plasmon excitation,

$$\text{Im} \left\{ \frac{-1}{\varepsilon(\omega, q)} \right\}_{pl} = g(\omega_0) \left. \frac{\pi}{|d\varepsilon_L^r(q, \omega; \omega_0)/d\omega_0|} \right|_{\varepsilon_L^r=0} \Theta \left[ q^-(\omega; \omega_0) - q \right], \quad (14)$$

where the single-valued  $\omega_0$  satisfies  $\varepsilon_L^r(q, \omega; \omega_0) = 0$ ; it is a numerical solution of the plasmon frequency at  $q = 0$  for the plasmon dispersion line that passes through the given  $(q, \omega)$ -point. The slope of the real part of the Lindhard dielectric function is

$$\begin{aligned} d\varepsilon_L^r(q, \omega; \omega_p)/d\omega_p \Big|_{\varepsilon_L^r=0} &= -\frac{2}{3\omega_p} \left\{ 2 + \frac{1}{\pi a_0 q} \frac{1}{Z} + \frac{1}{2\pi a_0 q} \frac{1}{Z} \right. \\ &\times \left. \left[ 2 - C_1 \ln \left| \frac{C_1 + 1}{C_1 - 1} \right| - C_2 \ln \left| \frac{C_2 + 1}{C_2 - 1} \right| + \frac{X}{4Z^2} \left( C_1 \ln \left| \frac{C_1 + 1}{C_1 - 1} \right| - C_2 \ln \left| \frac{C_2 + 1}{C_2 - 1} \right| \right) \right] \right\}, \quad (15) \end{aligned}$$

where  $C_1 = Z - X/4Z$  and  $C_2 = Z + X/4Z$ . The condition  $\Theta \left[ q^-(\omega; \omega_0) - q \right]$  requires the plasmon dispersion line for  $\varepsilon_L^r(q, \omega; \omega_0) = 0$  to be terminated at the left boundary of the area

for single electron excitation. For Al, the dominant intensity of  $\text{Im}\{-1/\varepsilon(q, \omega)\}_{pl}$  is around the bulk plasmon frequency, so  $\omega_0 \approx 15$  eV with certain expansion for the peak width. The total energy loss function is the summation of these two parts,

$$\text{Im}\left\{\frac{-1}{\varepsilon(q, \omega)}\right\} = \text{Im}\left\{\frac{-1}{\varepsilon(q, \omega)}\right\}_e + \text{Im}\left\{\frac{-1}{\varepsilon(q, \omega)}\right\}_{pl}. \quad (16)$$

On the other hand, by SPA the Lindhard energy loss function is simply given by

$$\text{Im}\left\{\frac{-1}{\varepsilon_L(q, \omega; \omega_p)}\right\} \approx \frac{\pi \omega_p^2}{2 \omega_q} \delta(\omega - \omega_q), \quad (17)$$

where the equation

$$\omega_q^2(\omega_p) = \omega_p^2 + \frac{1}{3} v_F^2(\omega_p) q^2 + (\hbar q^2 / 2m)^2 \quad (18)$$

defines the plasmon dispersion  $\omega_q$ , and  $v_F(\omega_p)$  is the Fermi velocity of an electron gas with the plasmon frequency  $\omega_p$ . The energy loss function then becomes

$$\text{Im}\left\{\frac{-1}{\varepsilon(q, \omega)}\right\} \approx \frac{\omega_0}{\omega_q} \text{Im}\left\{\frac{-1}{\varepsilon(\omega_0)}\right\}, \quad (19)$$

where  $\omega_0$  is the solution of equation  $\omega_q(q, \omega_0) = \omega$ . Basically,  $\omega_0$  is similarly obtained as in FPA except that now the dispersion equation is given explicitly. SPA is a good approximation for materials, such as transition and noble metals, that have broad optical energy loss functions in the loss energy region of  $10^0$ - $10^2$  eV, but not for free-electron-like materials for which a sharp plasmon peak dominates the optical energy loss function.

Fig. 2 shows a perspective view of the energy loss function obtained by FPA and SPA for Al and Cu. The optical data for Al and Cu are taken from (Shiles et al., 1980) and (Hagemann et al., 1974), respectively. The difference between the two methods is mostly significant at the low loss-energy and low momentum-transfer area (Fig. 2); for energy loss and momentum transfer higher than shown in Fig. 2, the energy loss function forms the Bethe ridge without apparent difference between FPA and SPA. For Al, the FPA energy loss function still has a limited but nonzero intensity for single particle excitation even for  $\hbar\omega < \hbar\omega_p$ . This is an important source for the creation of low energy secondary electrons in a MC simulation; the plasmon excitation intensity decays quickly when the dispersion enters into the single particle excitation region (Fig. 2(a)). SPA on the other hand completely ignores the single electron excitation. This missing contribution is compensated to the intensity of plasmon excitation whose dispersion line extends up to large  $q$  values while the ridge height decays very slowly (Fig. 2(b)). Therefore, there is no any energy loss when  $E < \hbar\omega_p$ . This becomes a serious problem for low energy electron inelastic scattering. For Cu, the differences still exist (Figs. 2(c)-(d)), but owing to the rather smooth and broad shape of the optical energy loss function the inelastic scattering probability obtained by the two methods for low losses and at low energies should be comparable.

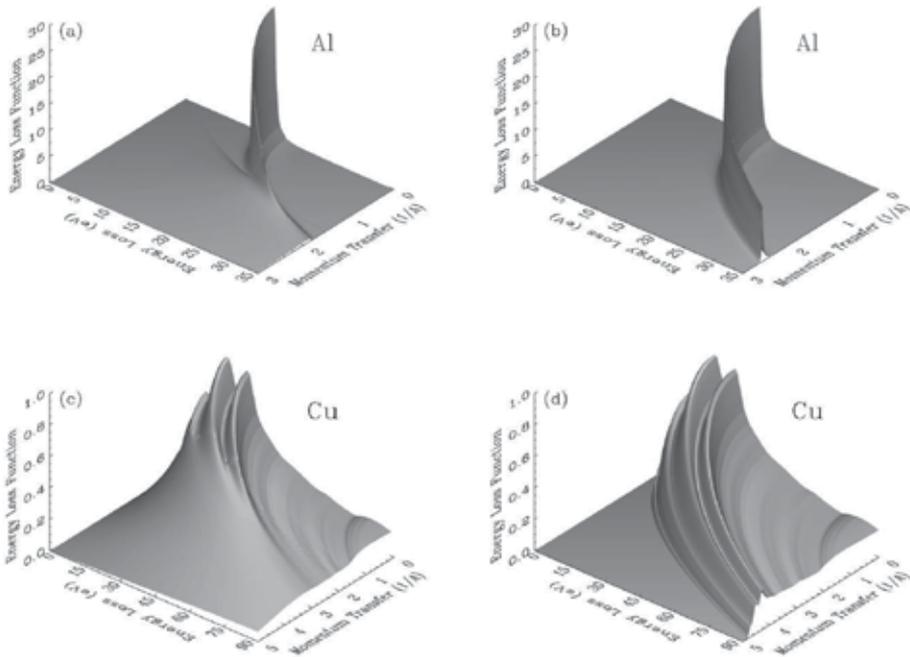


Fig. 2. Perspective plot of the energy loss function as a function of momentum transfer and energy loss, calculated by the full Penn algorithm (FPA) and single-pole approximation (SPA) for Al and Cu: (a) Al, FPA; (b) Al, SPA; (c) Cu, FPA; (d) Cu, SPA.

### 2.1.3 Electron cascading

MC simulation is a powerful technique for simulation of secondary electron emission phenomena and calculation of the related energy spectra and yields of secondary electrons (Shimizu & Ding, 1992; Ding & Shimizu, 1996; Ding et al., 2001; Ding et al., 2004a). As described above, the generation of SEs is a cascade process due to multiple inelastic scattering. The differential inverse inelastic mean free path given by Eq. (6) is used to describe the energy loss of moving electrons, as well as the cascade secondary electron generation.

The SPA was used in our previous simulation of SEs (Ding et al., 2001). After sampling the energy loss,  $\Delta E$ , for an inelastic scattering event, a SE is assumed to be excited from the Fermi sea by adding the loss energy  $\Delta E$  of the moving electron to the initial kinetic energy of the SE, with the excitation probability being proportional to a joint density of states of free electrons, i.e.  $p(E', \Delta E) \propto \sqrt{E'(E' + \Delta E)}$ , where  $E' < E_F$  is the energy of the Fermi sea electrons. This assumption follows Chung's work (Chung & Everhart, 1977) on plasmon damping. Under this procedure, with the differential energy loss cross section or the secondary electron excitation function determined using SPA, all SEs were assumed to be excited in the same way as through plasmon damping.

However, the treatment of secondary electron generation by FPA needs to consider two individual mechanisms as in the calculation of differential inverse inelastic mean free paths, i.e. single electron excitation and plasmon damping. After the momentum transfer  $\hbar\mathbf{q}$  and the energy transfer  $\hbar\omega$  are determined by a MC sampling procedure, we specify that single

electron excitation occurs if  $q^-(\omega; \omega_p) < q < q^+(\omega; \omega_p)$ , and plasmon excitation occurs if  $q < q^-(\omega; \omega_p)$ , where the value of  $\hbar\omega_p$  is determined from the Fermi energy  $E_F$  through the relation with electron density.

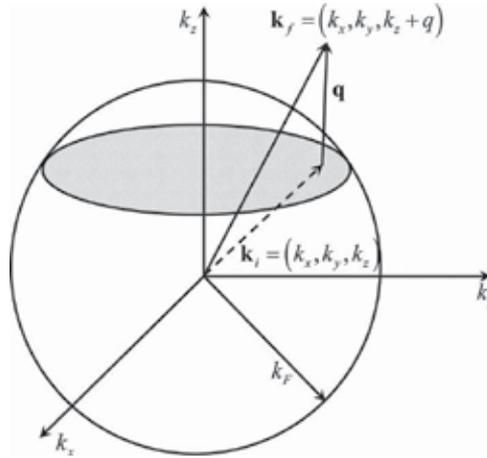


Fig. 3. Illustration of the condition for single electron excitation. The  $k_z$  axis is along the direction of  $\mathbf{q}$ . The shaded disk is the area that contains allowed momenta of excited electrons. An electron in the Fermi sphere at initial state of  $\mathbf{k}_i = (k_x, k_y, k_z)$  gains the loss energy  $\hbar\omega$  and momentum transfer  $\hbar\mathbf{q}$  from a scattering electron and transits to the final state of  $\mathbf{k}_f = (k_x, k_y, k_z + q)$  by energy and momentum conservations. Notice that if an electron excited from the interior of the allowed disk does not satisfy the condition that  $|\mathbf{k}_f| > k_F$ , the allowed disk becomes an annulus.

The secondary electron generation via plasmon damping can be treated exactly the same as in the previous model. For a single electron excitation, the probability distribution for exciting an electron of the momentum  $\mathbf{k}_i$  from the Fermi sea is given by (Ganachaud & Cailler, 1979a; Ganachaud & Cailler, 1979b),

$$p(|\mathbf{k}_i| < k_F; \mathbf{q}, \omega) = \int d\mathbf{k}_i \delta \left[ \hbar\omega - \frac{\hbar^2}{2m} (2\mathbf{k}_i \cdot \mathbf{q} + q^2) \right] \Theta(k_F - |\mathbf{k}_i|) \Theta(|\mathbf{k}_f| - k_F), \quad (20)$$

where  $\mathbf{k}_f = \mathbf{k}_i + \mathbf{q}$  is the momentum of the excited SE after the inelastic collision. As shown in Fig. 3, let the  $z$ -direction of the wave vector of the electron in the Fermi sea be the direction of the momentum transfer; by the energy and momentum conservations we have

$$\begin{cases} \frac{\hbar^2 \mathbf{k}_i^2}{2m} = \frac{\hbar^2}{2m} (k_x^2 + k_y^2 + k_z^2) = E' \\ \frac{\hbar^2 \mathbf{k}_f^2}{2m} = \frac{\hbar^2}{2m} [k_x^2 + k_y^2 + (k_z + q)^2] = E' + \hbar\omega \end{cases}, \quad (21)$$

or, equivalently, the components of  $\mathbf{k}_i = (k_x, k_y, k_z)$  have to satisfy two conditions: (1)  $k_x^2 + k_y^2 + k_z^2 < 2mE_F/\hbar^2$ ; (2)  $k_x^2 + k_y^2 + (k_z + q)^2 > 2mE_F/\hbar^2$ . The electrons that can be excited therefore lie in a disk of Fermi sphere defined by

$$k_z = (2m\omega - \hbar q^2) / 2\hbar q . \tag{22}$$

The momentum  $\mathbf{k}_i$  of the electron to be excited can be selected from the disk with two random numbers. This determines the momentum  $\mathbf{k}_f$  and the kinetic energy  $\hbar^2\mathbf{k}_f^2/2m$  of the secondary electron after excitation.

**2.1.4 General Monte Carlo simulation procedure**

Base on the knowledge of electron elastic and inelastic scattering and electron cascading described above, a general MC simulation procedure is given below (Ding, 1990; Ding & Shimizu, 1996).

The present MC simulation of electron trajectories penetrating a sample is based on a description of individual electron scattering processes, as schematically shown in Fig. 4. The problem is, then, reduced to the determination of values of physical quantities such as step length, scattering angle, energy loss, and so forth, in a particular scattering event. The MC technique basically chooses these values by random numbers according to respective cross sections. Given a probability distribution function  $P(x)$  for a variable  $x$ , we can derive a normalized accumulation function  $A(x)$ ,

$$A(x) = \int_{x_{\min}}^x P(x') dx' / \int_{x_{\min}}^{x_{\max}} P(x') dx' , \tag{23}$$

and determine a specific value of  $x$  from  $A(x) = R$  for a given value of uniform random number  $R \in [0,1]$ .

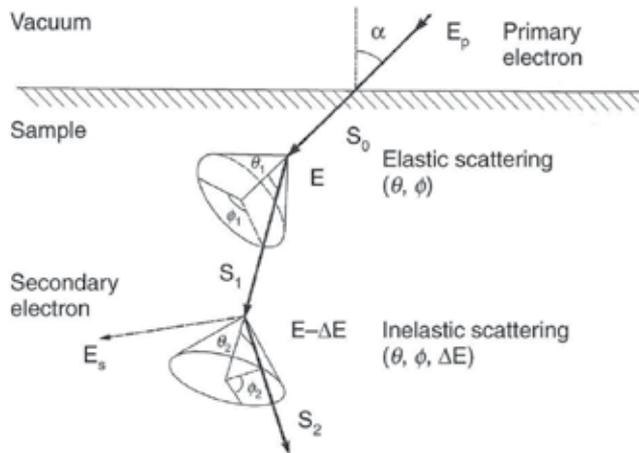


Fig. 4. Scheme showing the discrete model for Monte Carlo simulation of electron scattering.

Suppose that the step length,  $s$ , of a scattering electron between two successive collision events obeys the Poisson stochastic process with the probability distribution

$$P(s) = \lambda_m^{-1} e^{-s/\lambda_m} , \tag{24}$$

where  $\lambda_m$  is the total mean free path related to the corresponding elastic mean free path and the inelastic mean free path through

$$\lambda_m^{-1} = \lambda_e^{-1} + \lambda_{in}^{-1}. \quad (25)$$

An electron will then suffer a scattering event when it passes  $s$  selected by a random number  $R_1$  via

$$s = -\lambda_m \ln R_1. \quad (26)$$

Another random number,  $R_2$ , determines the type of individual scattering event followed after passing  $s$ : If

$$R_2 < \lambda_e^{-1} / \lambda_m^{-1}, \quad (27)$$

it is elastic, otherwise it is inelastic.

For elastic scattering, it is attributed to an atom of  $i$  th element if

$$\frac{\sum_{j=1}^{i-1} C_j^a / \lambda_e^j}{1/\lambda_e} < R_3 < \frac{\sum_{j=1}^i C_j^a / \lambda_e^j}{1/\lambda_e}, \quad (28)$$

where  $(\lambda_e^0)^{-1} = 0$ , holds. The angle of scattering is calculated by

$$R_4 = \int_0^\theta \left( \frac{d\sigma_e}{d\Omega} \right)_{\text{Mott}} \sin \theta' d\theta' / \int_0^\pi \left( \frac{d\sigma_e}{d\Omega} \right)_{\text{Mott}} \sin \theta' d\theta', \quad (29)$$

with the differential cross section for  $i$ th element, and the azimuthal angle is assumed to be isotropic,

$$\phi = 2\pi R_5. \quad (30)$$

Regarding inelastic scattering, the equation similar to Eq. (28) can be applied by replacing  $\lambda_e$  with  $\lambda_{in}$ , and  $\lambda_e^j$  with  $\lambda_{in}^j$ , for an alloy-like sample described by the sets of dielectric functions for each component, in the sense that it is not an atom but the constitutive electrons themselves that contribute to inelastic collision. For a compound-like sample, the dielectric function is given such as to treat the material as a whole, and it is therefore not necessary to specify a particular element. In the next step we determine the amount of energy loss  $\Delta E$  by the equation

$$R_6 = \int_0^{\Delta E} \frac{d\lambda_{in}^{-1}}{d(\Delta E')} d(\Delta E') / \int_0^{E-E_F} \frac{d\lambda_{in}^{-1}}{d(\Delta E')} d(\Delta E'), \quad (31)$$

It should be noted that the zero level of kinetic energy of electrons has been taken as the bottom of the valence band. The inelastic scattering angle is calculated by

$$R_7 = \int_0^\theta \frac{d^2 \lambda_{in}^{-1}}{d\Omega d(\Delta E)} \sin \theta' d\theta' / \int_0^\pi \frac{d^2 \lambda_{in}^{-1}}{d\Omega d(\Delta E)} \sin \theta' d\theta', \quad (32)$$

once the energy loss  $\Delta E$  is determined in a preceding stage by Eq. (31), where the double differential cross section with respect to energy loss and solid angle can be obtained from

Eq. (6) by transforming variable  $q$  to  $\theta$ . The energy and the momentum conservation give the relation

$$(\hbar q)^2/2m = 2E - \Delta E - 2\sqrt{E(E - \Delta E)} \cos \theta, \quad (33)$$

with which the change of variable can be made. Hence, we have

$$\frac{d^2 \lambda_{in}^{-1}}{d\Omega d(\Delta E)} = \frac{1}{(\pi a_0 e)^2 E} \text{Im} \left\{ \frac{-1}{\varepsilon(q, \omega)} \right\} \frac{1}{q^2} \sqrt{E(E - \Delta E)}. \quad (34)$$

It can be shown that the angular distribution is proportional to  $\theta^{-2}$  for small  $\theta$  and the forward scattering is strongly preferred. The azimuthal angle in inelastic scattering is also determined by Eq. (30).

Followed by an electron inelastic scattering event, in which the energy loss  $\Delta E$ , momentum transfer  $\hbar q$  and coordinates (the position where inelastic scattering happens) are suitably determined, a SE would be generated; its initial energy  $E_2$  (the kinetic energy equals to  $E_2 + \Delta E$  after excitation) is determined from

$$R_8 = \frac{\int_0^{E_2} \sqrt{E'(E' + \Delta E)} dE'}{\int_0^{E_F} \sqrt{E'(E' + \Delta E)} dE'}, \quad (35)$$

in SPA, or, for plasmon excitation in FPA. The polar and azimuthal angles of the cascade electron are decided from momentum conservation as

$$\sin \theta_2 = \cos \theta, \quad \phi_2 = \pi + \phi. \quad (36)$$

For single electron excitation in FPA, by sampling a point in a disk/annulus (Eq. (22)), the energy and direction of the SE can be also easily decided by Eq. (21). The coordinates of the SE is set as that at the scattering point. As the energy, direction and coordinates of the SE are determined, they are stored in memories of a computer. After finishing the tracing of the primary electron all the information of the cascade electron stored are recalled and the trajectory is simulated in a same way as for a primary electron.

The scattering angle in Eq. (29) and Eq. (32) and the azimuthal angle in Eq. (30) are given in a coordinate system moving with the tracing electron (Fig. 4). When these angles are transformed into a coordinate system fixed in the sample, the position at the next scattering point can be determined. Repeating above procedures, we get a trajectory of the penetrating electron which either terminates in the sample when its kinetic energy falls below a cut-off energy  $E_c$  or escapes from the surface.

It should be noted that in the simulation, the reference energies (bottom of valence band or vacuum level) depends on electron location. For a primary electron penetrating into the surface from vacuum the energy of the electron before first inelastic collision is given by

$$E = E_p + U_0, \quad (37)$$

where the primary energy,  $E_p$  is measured from the vacuum level and  $E$  from the bottom of valence band. The inner potential,  $U_0$ , is approximated by the sum of Fermi energy and work function,

$$U_0 = E_F + W_F. \quad (38)$$

In the case of electron ejection into the vacuum from the sample, the energy measured from the vacuum level is

$$E = E' - U_0, \quad (39)$$

and the ejection angle,  $\vartheta$ , measured from the surface normal, is found by the momentum conservation parallel to surface (like the light deflection at the interface of media),

$$\sqrt{E} \sin \vartheta = \sqrt{E'} \sin \vartheta', \quad (40)$$

where the superscript denotes the corresponding quantities inside the sample. It should be noted that the surface barrier inhibits an electron from escaping from the surface into vacuum with an angle  $\vartheta' > \vartheta_c$ , where

$$\vartheta_c = \sin^{-1} \sqrt{1 - \frac{U_0}{E'}} = \cos^{-1} \sqrt{\frac{U_0}{E'}}. \quad (41)$$

The quantum mechanical representation of the transmission coefficient is given by (Cohen-Tannoudji et al., 1977)

$$T_q(E', \vartheta') = \begin{cases} \frac{4\sqrt{1 - U_0/E'} \cos^2 \vartheta'}{\left[1 + \sqrt{1 - U_0/E'} \cos^2 \vartheta'\right]^2}, & \text{if } E' \cos^2 \vartheta' > U_0; \\ 0, & \text{otherwise.} \end{cases} \quad (42)$$

By using another random number  $R_8$ , whether the electron could emit or not can be decided as

$$\begin{cases} \text{emitted,} & \text{if } R_8 < T_q(E', \vartheta'); \\ \text{absorbed,} & \text{otherwise.} \end{cases} \quad (43)$$

## 2.2 Geometry modeling of complex structure

The request of simulation for inhomogeneous sample with complex geometry structure demands an efficient structure modeling to be included in a MC simulation. There are two questions to be settled, i.e., how to construct a complex geometry and how to correct the flight path between two successive collisions.

To solve the first question, the CGS modeling (Ding & Li, 2005; Li & Ding, 2005) and finite element triangle mesh modeling (Ding & Wang, unpublished; Wang, 2006; Li et al., 2008; Li, 2009) have been bring forward. The detail will be introduced below.

To the second question, considering an inhomogeneous sample formed of many different spatial zones, each of which is, however, homogeneous in atomic composition so that the total cross section  $\sigma = \lambda_m^{-1}$  is only a constant for an electron traveling within a specific zone. Therefore, the case of continuous change of atomic composition will not be considered here. A problem in sampling by Eq. (26) occurs when an electron crossing the interface of zones with different  $\sigma$  values. Let  $T_i$  denote the segment of electron step length  $s$  within the  $i$ -th

zone of scattering cross section  $\sigma_i$ , the Eq. (26) can then be simplified to a sum of a discrete sequence,  $\sigma_i T_i$ , as

$$\sum_i \sigma_i T_i = -\ln R \quad \text{and} \quad s = \sum_i T_i. \quad (44)$$

This differs from the conventional sampling that the step length is only associated with the scattering cross section at starting location of the flight step. This correction to the step length is expected to be important for specimen containing nanoscale structures so that the electron scattering mean free path, that is, the inverse of scattering cross section, is comparable with or even larger than the structure feature size. Then the question reduces to obtaining the partial distance  $T_i$  for an electron flying over the different specimen zone and the corresponding sequential points intersecting at the zone surfaces. The detail of the sampling process is performed by using a ray-tracing technique (Ding & Li, 2005), which will be shown below. Besides the path length sampling, an electron would be refracted at the boundary interface. A treatment of the refraction will also be explained.

### 2.2.1 CSG modeling

The geometric structure of the specimen should be specified at first before the simulation. By CSG modeling, a complex geometric structure can be constructed with some simple and basic shapes which can be analytically described with a few parameters (Ding & Li, 2005; Li & Ding, 2005). The blocks enclosed by these shapes contain either the different materials or even simply empty ( $\sigma = 0$ ), as illustrated by Fig. 5. For each electron trajectory step, one has to compute the intersecting points with every possible shape. Because a MC simulation requires tracking a large number of electron trajectories, the computation for judging intersecting points is the most time consuming in a simulation. Therefore, the structure construction technique needs to be efficient for the calculation. Here we choose a half-infinite space with flat top surface as the specimen basis on which a complex structure is constructed with some basic shapes. The basic shapes include sphere, ellipsoid, cylinder, cone, cube, tetrahedron, polyhedron, and so forth, which can be easily and analytically defined. Each kind of basic shape may be implemented into a subroutine to allow for the calculation of intersecting points efficiently.

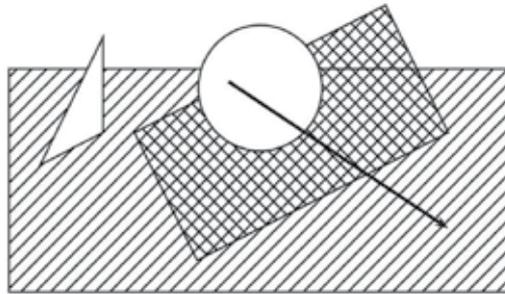


Fig. 5. The sketch of constructing specimen geometric structure by CSG modeling.

It is obvious that, with the limited number of parameters, one cannot in practice build an arbitrary complex structure shape. Fortunately, most of the geometric structures observed with SEM can be modeled by the present algorithm. In particular, the use of empty blocks

enclosed by the basic shapes allows for some special structures to be constructed, for example, porous material. Generally, for each shape used in the construction one has to transform the orientation of the basic shape in a coordinate system fixed with the sample to another one fixed with the shape so that the standard method for solving the simple form is applicable. More detail description of this procedure is given by Li & Ding's work (Li & Ding, 2005).

### a. Judging intersecting and sorting

A series of subroutines for determining the intersecting points of a ray with the basic shapes were developed. Because an arbitrary specimen structure is made up of some basic shapes which can be spatially located in any position, we have to judge intersecting points of one electron trajectory step, viewed as a ray, with every basic shapes specified. In this way all the distance  $T'$  pairs can be obtained, where  $T'$  denotes the distance between the starting point and a particular intersecting point, for an electron trajectory intersecting with each shape (one  $T'$  for incoming and another for outgoing) along the electron moving direction of the electron motion. A series of subroutines is used to calculate the  $T'$  pairs for each kind of shape so that the program could be modularized; its advantage is obviously that it is easy to expand of the program when including other newly added basic shape types (Li & Ding, 2005).

Having obtained all the intersecting points and the  $T'$  pairs, we have to calculate the distance  $T$  for an electron moving within a particular material block. To obtain the distance between the adjacent intersecting points along the electron path, the right  $T'$ -sequence should first be derived. A simple sort subroutine based on Shell-sort (which, as one of the oldest sorting algorithms, is fast and easily implemented) can do this well (Knuth, 1998). A 2-D array is then used to save the Tsequence and the index for the material to label the corresponding scattering cross sections. The correct  $\sigma_i$  is thus determined for the material between the  $(i-1)$ -th and  $i$ -th intersecting points. This procedure is illustrated by Fig. 6.

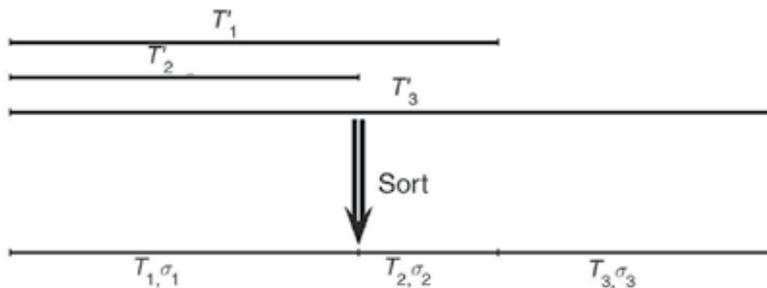


Fig. 6. Schematic diagram for sorting the intersection distances and the corresponding cross sections.

Regarding the efficiency of the above procedure, it is fast for the small number of shapes considered. Of course, one wishes to minimize judging the procedure steps when lots of shapes are involved in the structure building. Ideally, if we could know the order of rays intersecting with the basic shapes, we would only need to obtain the intersecting points in sequence until Eq. (44) was satisfied, so that only several related but not all shapes would be considered in the calculation. However, before obtaining the step length we cannot know the order. The above-mentioned method is one solution to this self-consistent problem.

For a large number of building shapes there are two ways of minimizing the necessary number of shapes involved in each step of the judging procedure. The first method is to construct a tree-shaped structure with which all the neighbors of a particular shape are specified. Then only the neighbors of the shape where an electron step is crossing are considered, from among which one correct shape is chosen and the intersecting points as well as the distance  $T$  are calculated. This procedure step continues until Eq. (44) is satisfied. Because in each procedure step we have to consider all the neighbors of the shape varying step by step, the overall computation for electron flight step length may still be inefficient. Another method considers the geometric structure constructed by dividing the whole space into small cubes. Although the neighbor judging is easier, the total computation time is also significantly higher because of the need to count many small cubic blocks.

### b. Calculation of path length

An electron flies in the specimen from the old scattering position  $\mathbf{x}_0$  to the new position  $\mathbf{x}$  over a distance  $s$  along the moving direction,

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{v}s. \quad (45)$$

The flight path vector  $\mathbf{v}s$  is treated as a ray in the judging procedure, where  $\mathbf{v}$  is the unit vector of velocity direction. The valid  $T$  in  $s = \sum_i T_i$  that represents the distances between intersecting points in Eq. (44) must be positive.  $T \leq 0$  can be excluded by the sorting procedure for obtaining the correct  $T$ -sequence.

The procedure for obtaining flight step length by Eq. (44) is as follows. First, a variable  $C$  is used to apply a do-loop to calculate

$$C = C + \sigma_i T_i, \quad i = 1, 2, \dots, m_c \quad (46)$$

where  $m_c$  is the number of intersecting points. If the condition  $C \geq -\ln R$  is satisfied at  $i = m$ , this means that the electron flight terminal will be a fall in the  $m$ -th section. The do-loop should be stopped at  $m$  and the step length is then given by

$$s = \sum_{i=1}^{m-1} T_i + \left( -\ln R - \sum_{i=1}^{m-1} \sigma_i T_i \right) / \sigma_m; \quad (47)$$

otherwise, there is no  $i$  for the condition being satisfied. This means the step length is longer than  $\sum_i T_i$ ; so some additional conditions should be used to decide  $s$ . There are two cases: (1) an electron is moving toward the vacuum, that is,  $v_3 < 0$  or  $z < 0$  and  $v_3 = 0$  and  $z = 0$ . The electron then escapes from the specimen and we can simply take  $s = \sum_{i=1}^{m_c} T_i$ . Another subroutine is used to determine whether the electron has enough energy to overcome the surface barrier; (2) an electron is moving toward the basis, that is,  $v_3 > 0$  or  $z > 0$  and  $v_3 = 0$  and  $z > 0$ . The electron is scattered in the basis, so the step length is

$$s = \sum_{i=1}^{m_c} T_i + \left( -\ln R - \sum_{i=1}^{m_c} \sigma_i T_i \right) / \sigma_0, \quad (48)$$

where  $\sigma_0$  denotes scattering cross section of the basis.

### 2.2.2 Finite element triangle mesh modeling

For a further study of the sample having arbitrary periphery, e.g. a rough surface, which is hard to be described analytically, a finite element triangle mesh, which is frequently used in computer graphics, has been introduced in the complex geometry modeling (Ding & Wang, unpublished; Wang, 2006; Li et al., 2008; Li, 2009). The principle of this method is to approximate a sample surface by using finite triangles. Obviously, more triangles are used more precise description of the sample surface is attainable. A schematic diagram for constructing complex sample geometry is shown in Fig. 7 (a)-(c). The space points are firstly selected to lie on the sample surface. By connecting the neighboring points in a certain order, as shown in Fig. 7 (b)-(c), a finite element triangle mesh for modeling of a sample surface with arbitrary geometry can be obtained. For example, in Fig. 7 (d), a finite element triangle mesh for a helix is shown.

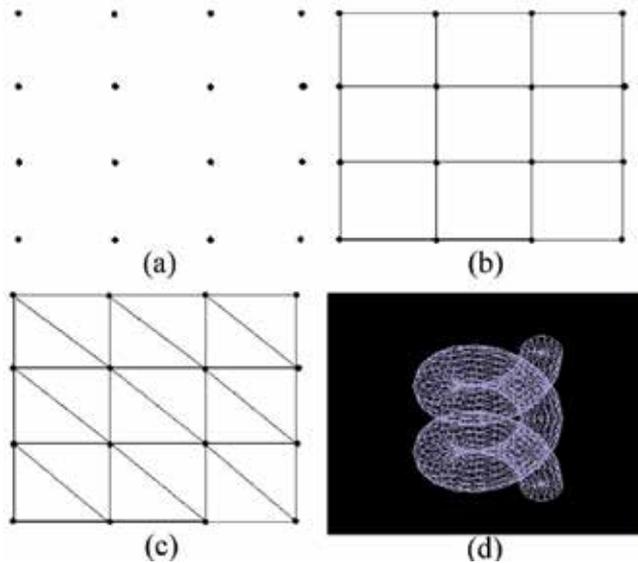


Fig. 7. Schematic diagram for constructing complex sample geometry. (a) Selection of the space points on sample surface. (b),(c) Connection of the neighboring point for obtaining the triangles. (d) The finite element triangle mesh for a helix.

In a MC simulation by using finite element triangle mesh modeling, it is hard to know whether an electron is located in the sample or not just by solving equations like in the CSG modeling. So, a tag should be added to the electron firstly to denote if it is in the sample or not. For example,

$$\xi = \begin{cases} 0, & \text{outside the sample;} \\ 1, & \text{in the sample.} \end{cases} \quad (49)$$

The tag would be changed when the electron trajectory passes through any triangle in the mesh. Judging intersection of the electron trajectory is just like that in CSG modeling. From present spatial location of an electron, a straight line which goes along the electron moving direction, is drawn. The line will encounter triangle(s) and be divided into several segments

by the intersection point, and each section has its tag to denote its position (in or out of the sample). Then a similar step like in CSG modeling can be used to calculate the path length.

### a. Space subdivision method

Although the basic algorithm for path length calculation is built, an improvement is still necessary for the simulation because the judging intersection of a trajectory with the triangles is quite time consuming. Here the space subdivision method is developed to accelerate the calculation. The scheme is shown by Fig. 8. After dividing the space into a cubic lattice containing the finite element triangle mesh (Fig. 8(a)), triangle slices lie in some cuboids (Fig. 8(b)). The intersection judging can be performed to be limited to within a subcuboid instead the whole space. To prepare the intersection judging, the statistics of the triangles contained in each subcuboid should be figured out at first. Starting from the present position of an electron, the first subcuboid on its path of the trajectory to be traveled can be determined, as well as  $C$  (Eq. (46)) in the first subcuboid. If  $C \geq -\ln R$  is satisfied, the path length is decided by Eq. (47). Otherwise, by using the straight line drawing algorithm (Cleary & Wyvill, 1988), the next subcuboid in the path of the trajectory can be found (Fig. 8(c)), and the same routine as in the first subcuboid is applied. After this do-loop is performed for subcuboids one by one, the path length is finally obtained; otherwise, the electron goes out of the sample.

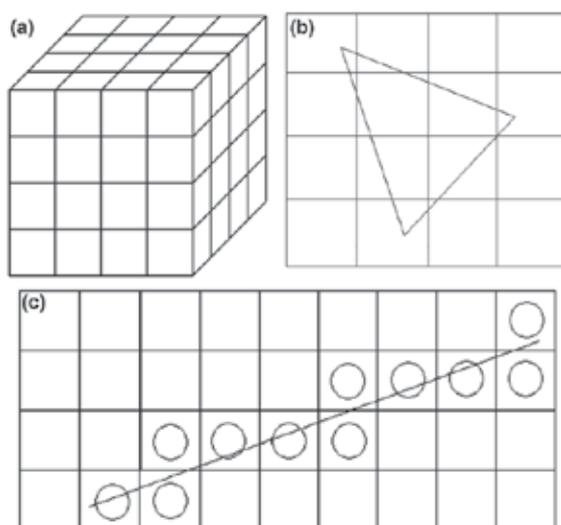


Fig. 8. Schematic diagram for space subdivision method. (a) Dividing space into cuboids. (b) Statistics of triangles lying in each cuboid. (c) Judging the subcuboids in which a trajectory is passed through.

### b. Treatment of refraction

When an electron penetrates through a boundary interface, the refraction of electron moving direction should be considered. The refraction at a surface is treated as follows: First, the direction of an electron in the sample coordinate system of axes  $(x, y, z)$  is transformed to the new coordinate system of axes  $(x', y', z')$  with the  $z'$ -axis normal to a triangulated plane (Fig. 9).

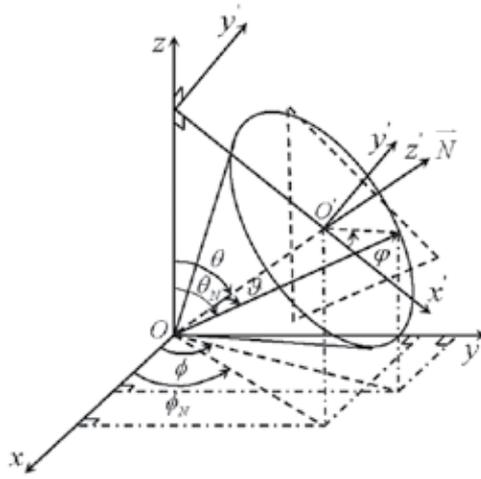


Fig. 9. The schematic diagram of coordinate system transformation for the directional vector  $\bar{N}$  of an electron. The old frame of axes  $(x, y, z)$  is fixed at the sample, and the new frame of axes  $(x', y', z')$  is at a local triangulated plane with the  $z'$ -axis normal to the plane. Here,  $\bar{N}$  stands for the normal vector of a local triangulated plane that the electron enters into.

The transform relation is,

$$\begin{pmatrix} v_{x'} \\ v_{y'} \\ v_{z'} \end{pmatrix} = \begin{pmatrix} \cos \theta_N \cos \phi_N & \cos \theta_N \sin \phi_N & -\sin \theta_N \\ -\sin \phi_N & \cos \phi_N & 0 \\ \sin \theta_N \cos \phi_N & \sin \theta_N \sin \phi_N & \cos \theta_N \end{pmatrix} \begin{pmatrix} u_x \\ u_y \\ u_z \end{pmatrix}, \quad (50)$$

where  $(v_{x'}, v_{y'}, v_{z'}) = (\sin \vartheta \cos \phi, \sin \vartheta \sin \phi, \cos \vartheta)$  is the unit vector of velocity in the new frame of axes and  $(u_x, u_y, u_z) = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)$  in the old frame of axes.  $\theta_N$  and  $\phi_N$  are the polar and azimuthal angles of the normal vector  $\bar{N}$  of the local triangulated plane, respectively. Second, when an electron escapes from solid into the vacuum a quantum mechanical transmission coefficient (Cohen-Tannoudji et al., 1977), i.e. Eq. (42), is applied to the electron fate as being absorbed or emitted.

Third, the direction of electron will be changed according to the refraction law,  $\sqrt{E_1} \sin \vartheta_1 = \sqrt{E_2} \sin \vartheta_2$ , where  $E_1$  and  $E_2$  are the electron energies inside and outside the solid, respectively;  $\vartheta_1$  and  $\vartheta_2$  are the polar angles of the unit vector inside and outside the solid, respectively. Because the electron kinetic energy inside a material is referenced to the top of conduction band, the energy changes by the inner potential when an electron passes through a surface/vacuum interface, i.e.  $E_2 = E_1 - U_0$ , where  $U_0$  is defined in Eq. (38). Thus,

the unit vector is changed as,  $(v'_{x'}, v'_{y'}, v'_{z'}) = (v_x S, v_y S, v_z \sqrt{1 - S^2(1 - v_z^2)})$ , where

$S = \sqrt{E_2/E_1}$ . When the polar angle  $\vartheta_1$  is larger than the critical angle,  $\vartheta_C = \sin^{-1} \sqrt{1 - U_0/E_1}$ ,

the electron may be reflected by the surface,  $(v'_{x'}, v'_{y'}, v'_{z'}) = (v_{x'}, v_{y'}, -v_{z'})$ . Fourth, an inverse transformation to the old frame of axes should be done after changing the direction by refraction,

$$\begin{pmatrix} u'_x \\ u'_y \\ u'_z \end{pmatrix} = \begin{pmatrix} \cos\theta_N \cos\phi_N & -\sin\phi_N & \sin\theta_N \cos\phi_N \\ \cos\theta_N \sin\phi_N & \cos\phi_N & \sin\theta_N \sin\phi_N \\ -\sin\theta_N & 0 & \cos\theta_N \end{pmatrix} \begin{pmatrix} v'_{x'} \\ v'_{y'} \\ v'_{z'} \end{pmatrix}. \quad (51)$$

### 3. Simulation of CD-SEM images for critical dimension nanometrology

In integrated circuit industry, new material and new device structures will probably allow MOS devices to remain competitive near-term years in spite of Moore's Law. Metrology of the electronic material and devices is crucial to the research and development of future semiconductor industry. Especially, the most fundamental one is that of dimension metrology, such as critical dimension (CD) and film thickness (Vogel, 2007). According to the International Technology Roadmap for Semiconductors (ITRS, 2007), the dimension of transistor gates has been decreased to about 30 nm with the accuracy lower than 1 nm by 2007. Thus, strict dimension control is an extremely urgent task in near-term years.

Many measurement methods are being developed to meet the case, such as, scatterometry, atomic force microscopy, transmission electron microscopy and SEM (Maeda et al., 2008). Among them, SEM is widely used as a standard tool for the linewidth measurement and CD metrology for its high resolution and high efficiency. A specialized length measuring instrument named critical dimension scanning electron microscope (CD-SEM) has thus been developed and widely used for the dimension metrology (Wang et al., 2007). The top-down mode in CD-SEM is mostly employed inline or offline for it is a nondestructive and great throughput examination.

However, there still remains a problem in accurate dimension metrology and linewidth measurement because of the edge effect in SE images. The edge effect can induce an important feature in SE line-scans, the bloom, which gives rise to the contrast that distinguishes the edge from the rest of the sample. The finite width of the bloom may lead to tens of nanometers of ambiguity in the edge position (Villarrubia et al., 2004; Tanaka et al., 2007). Moreover, the extent of the edge effect is the feature size dependent. When the line/feature size is decreased to several tens nanometers, the bias due to the edge effect would become more obvious. A lot of works including both experimental study (Bunday et al., 2007; Choi et al., 2006; Jones et al., 2003; Kawada et al., 2003; Maeda et al., 2008; Matsumoto et al., 2006; Morokuma et al., 2004; Novikov et al., 2007; Rice et al., 2006; Shishido et al., 2002; Tanaka et al., 2003; Tanaka et al., 2004; Tanaka et al., 2005; Tanaka et al., 2007; Tanaka et al., 2008a; Tanaka et al., 2008b; Wang et al., 2007; Yamane & Hirano, 2005) and theoretical investigation (Abe et al., 2007; Babin et al., 2008a; Babin et al., 2008b; Bunday & Allgair, 2006; Dersch et al., 2005; Frase & Hqßler-Grohne, 2005; Frase et al., 2007a; Frase et al., 2007b; Gorelikov et al., 2005; Villarrubia et al., 2004; Villarrubia et al., 2005a; Villarrubia et al., 2005b) with MC simulation methods have been done, aiming at accurate estimation of the CD values with CD-SEM. However, when the dimension decreases to tens nanometers most of the experiential methods, such as, the maximum derivative method, the regression to baseline method and the sigmoidal fit method, face different difficulties (ITRS, 2007; Villarrubia et al., 2005a). A reasonable algorithm is urgently needed for the linewidth metrology of nanometer systems. Several algorithms (Bunday & Allgair, 2006; Frase & Hqßler-Grohne, 2005; Frase et al., 2007a; Morokuma et al., 2004; Novikov et al., 2007; Shishido et al., 2002; Tanaka et al., 2003; Villarrubia et al., 2004; Villarrubia et al., 2005a; Villarrubia et al., 2005b) by MC methods for CD linewidth determination have been brought up to meet the limitation of the traditional experiential algorithms.

Recently, several new algorithms have been published. Novikov et al. (Novikov et al., 2007) have described a method for linear measurement in the nanometer range by taking into account the relationship between the specific probe positions and the SEM signal sharp kink points. The method is quite simple and suitable for large linewidth and sidewall angle; but, it would have large error when linewidth and/or sidewall angle are small because the kink points can not be distinguished clearly in these cases. Frase et al. have presented an exponential distribution operators method (Frase & Hqßler-Grohne, 2005; Frase et al., 2007a) with which the SEM intensity profile is modeled by a piecewise-defined continuous function that approximates to the measured intensity profile extracted from an image by means of a least-squares fit. The algorithm is tested by a series of MC simulations which fit well with the simulation results for large dimensions (>100 nm) and large sidewall angles (not close to vertical). However, the error will be large also for smaller dimensions and sidewall angles. Tanaka et al. have developed a multiple parameter profile characterization algorithm (Morokuma et al., 2004; Shishido et al., 2002; Tanaka et al., 2003) based on MC simulated and experimental results, which partitions the SEM image signal into the sidewall and footing based on the first deviation of the measured signal level. It applies to top-down SEM images and no throughput loss will be incurred; it is shown to have a 3-sigma accuracy of  $\pm 0.9^\circ$  for sidewall angle deviating by more than  $2^\circ$ . The limit of the algorithm is the effective range of the sidewall angle estimation, and, the error will also increase when the linewidth is reduced. Villarrubia et al. have introduced a model-based library (MBL) method (Villarrubia et al., 2004; Villarrubia et al., 2005a; Villarrubia et al., 2005b) by which a library of the MC simulation results for various parameters spanning the process space of interest is constructed. Dmitry et al. have also used an off-line generated MC simulation library and fitted the measured intensity profiles of the critical shape metrology (Gorelikov et al., 2005). In principle, this library method would be the most accurate one for it can directly relate the measured signal profiles with the modeled geometry. However, it would need to put huge effort in order to obtain such a calculation library for lots of different experimental parameters, and it is difficult to differentiate cases with the similar shape of SE line-scan profiles.

Frase et al. have recently reviewed the fundamentals, special performance features and applications of existing SEM image contrast simulation packages based on MC methods (Frase et al., 2009). It is considered that the MBL method should be the most accurate algorithm for its sound physical foundation and one-to-one relationship between the image intensity profile and the geometry models (Villarrubia et al., 2005a). The model-based method by using a MC simulation has been proved to be an excellent approach to determine specimen geometric parameters by comparing the measured SEM image data directly with the model input data for calculation. For this purpose we have to deal with three elements involved in a simulation, i.e. the sophisticated physical model, the universal geometric structure model and efficient simulation algorithm. For SEM imaging simulation the physical modeling of SE signal generation in different materials plays a critical role. The influence of different MC models to the linewidth determination has been carefully carried out recently (Villarrubia & Ding, 2009). For a geometric structure modeling of a realistic trapezoidal line shape, many parameters should be taken into account, such as, width, height, foot/corner rounding, sidewall angle and roughness, etc. Apart from these, other influencing factors, such as electron beam condition (primary energy, probe size and incidence angle), material properties (composition and distribution), SE signal detection (detector properties and electric field) and others (charging effects, noise and etc.) should be

considered to meet the accurate linewidth control (Frase & Haßler-Grohne, 2005). In order to build such a library an efficient simulation program should be constructed in order to save calculation time for a mass of practical conditions.

Among them, however, the influence of the side wall roughness, the line edge roughness (LER) or the line width roughness (LWR), to CDs has rarely been included in a theoretical calculation because of the complexity of constructing a reasonable rough surface model. The LER or the LWR can degrade resolution and linewidth accuracy (Yoshimura et al., 1993) and cause fluctuation of transistor performance (Asenov et al., 2003; Croon et al., 2002; Diaz et al., 2001; Ercken et al., 2002; Hamadeh et al., 2006; Kaya et al., 2001; Kim et al., 2004a; Kim et al., 2004b; Linton et al., 1999; Linton et al., 2002; Oldiges et al., 2000; Xiong & Bokor, 2002; Xiong & Bokor, 2004; Xiong et al., 2004; Yamaguchi et al., 2003; Yamaguchi et al., 2004). It becomes a critical issue when the CDs for semiconductor devices shrink into few tens nanometers (ITRS, 2007; Gwyn et al., 2003) because the roughness on the edge of the line does not scale with the linewidth (Asenov et al., 2003). Though many efforts have been done to estimate LER/LWR with top-view SEM images (Braun, 2005; Foucher et al., 2006) and especially by comparing with CD-AFM (Foucher et al., 2006), their dependence on different CD-SEM experimental conditions is still not quite clear because the roughness is a complex parameter to metrology. But, on the other side, in a MC simulation one can in principle construct a rough surface structure model with exact defined values of roughness parameters. The image contrast simulation enables us to establish a relation of SEM image contrast with structure model directly; such a relationship can be used for a quantitative estimation of the influence of roughness on the linewidth measurement. For this purpose we shall introduce a universal rough edge structure modeling into a MC simulation for the first time.

In this section, a MC simulation program for modeling of linewidth measurement has been developed to solve these problems (Ding & Shimizu, 1996; Ding & Li, 2005; Ding & Wang, unpublished; Li & Ding, 2005; Li et al., 2008; Mao et al., 2008; Shimizu & Ding, 1992; Yue et al., 2005). The main topic here is to apply the MC method to investigate in detail the influence of various factors to the contrast of SEM line-scan and image. The study gives a further insight into the new algorithm for CD metrology/linewidth determination. This MC simulation is mainly based on the up-to-date electron scattering model (Shimizu & Ding, 1992; Ding & Shimizu, 1996; Mao et al., 2008), a universal geometric structure model (Ding & Li, 2005; Ding & Wang, unpublished; Li & Ding, 2005; Li et al., 2008; Yue et al., 2005) and the message passing interface (MPI) program with accelerated algorithm (Ding & Wang, unpublished; Li et al., 2008), as mentioned in Sec. 2 and the following. A systematic calculation for different parameters has been done, advancing a step towards building a MBL library.

### 3.1 Simulation method and surface roughness model

Villarrubia and Ding (Villarrubia & Ding, 2009) have compared eight MC physical models based on phenomenological fitting measured parameters, a binary scattering model and a dielectric function approach. They concluded that, CD linewidths estimated by these models agree to each other within  $\pm 2.0$  nm on silicon and  $\pm 2.6$  nm on copper in 95% of comparisons with electron landing energy, beam width, and other parameters typical of those used in industrial CD measurements. In Ref. (Frase et al., 2009), Frase et al. also reviewed the physics of probe-sample interaction and modeling, and existing MC simulation programs. Both of them found that, the MC model based on the Mott's cross-section for elastic

scattering and a dielectric function approach for inelastic scattering would achieve the most agreeable result comparing with experiment. Furthermore, this MC model enables a more detailed treatment of the various physical processes involved, such as, the cascade SE generation.

In this respect, the MC model (Li et al., 2008) used here for SEM image simulation has been improved as shown in Sec. 2. Additionally, a rough surface geometry model is introduced to construct the sample surface, together with using of a ray-tracing technique (Li & Ding, 2005) in the calculation procedure of electron step length. Appropriate boundary correction (Yue et al., 2005) had also been considered in order to treat the reflection/refraction of low energy SEs when they pass through an interface separating different materials.

### 3.1.1 Electron scattering model

The MC simulation of electron scattering process bases on the tracing of incident electron trajectories made of joining of randomly sampled electron scattering events as well as that of generated SE (Ding & Shimizu, 1996; Shimizu & Ding, 1992). For the treatment of electron elastic scattering (also can be found in Section 2), the Mott's cross section (Mott, 1929) with the Thomas-Fermi-Dirac atomic potential (Bonham & Strand, 1963) is employed. As for electron inelastic scattering, we use a dielectric function formalism which handles as well the SE production in an electron inelastic scattering event. The FPA has been used in this work to calculate energy loss function (Mao et al., 2008), without introducing SPA (Ding & Shimizu, 1996; Penn, 1987). The compiled experimental data on the optical constants (Palik, 1991) were used. Secondary electron excitation process is divided into two individual parts in the calculation (Mao et al., 2008), i.e. the single electron excitation and the excitation via plasmon decay in different areas of the momentum transfer- and energy loss-plane. This consideration has been proven to be more exact than the previous model (Ding & Shimizu, 1996) in deriving secondary electron energy distribution particularly for those materials that having a strong and sharp plasmon energy loss peak in its optical energy loss function, such as, Si and Al.

As pointed out by Villarrubia & Ding (Villarrubia & Ding, 2009), the physical model itself could introduce error more or less, in spite of using the most sounded physics. Here, in order to stand out the influence of other factors to the CDs, we would neglect the uncertain error due to the physical model here.

### 3.1.2 Sample structure construction

Modeling of 3D structures is the basic factor in contemporary CD metrology, thus, different kinds of modeling have been introduced (Frase et al., 2009). The reasonability for illustration of 3D specimen structures and the optimization of simulation are certainly the two most aspects for a 3D geometric structure model.

The geometric structure used in this simulation is mainly based on our primary works (Ding & Li, 2005; Ding & Wang, unpublished; Li & Ding, 2005; Li et al., 2008; Yue et al., 2005), as illustrated in Sec. 2. For describing a geometrical structure, two main approaches have been applied; one is the CSG (Ding & Li, 2005; Li & Ding, 2005; Yue et al., 2005) and the other uses a finite element triangle mesh to approximate the sample surface (Ding & Wang, unpublished; Li et al., 2008). Furthermore, for the surface roughness structure, the finite element triangle mesh model is very suitable to simulate the line edge roughness of the wafer gate. The surface roughness can be parameterized by the amplitude  $3\sigma$  and the density described by the interval  $a$  of rough peaks.

A Gaussian function is a good approximation for characterization of gate length fluctuation or linewidth roughness (Bunday et al., 2004; Kim et al., 2004a; Xiong et al., 2004) and is usually used to model the rough effects. Therefore, here we represent the random surface roughness by introducing two parameters: the  $3\sigma$  deviation of a Gaussian function describes the amplitude fluctuation of rough peaks, and, the mesh interval,  $a$ , describes the density of rough peaks. By varying the two parameters we can obtain a variety of rough surfaces or rough edges. Specific sampling of vertical coordinate above the plane,  $z$ , which satisfies the statistical distribution,

$$f(z) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-z^2/2\sigma^2), \quad (52)$$

for each grid point is firstly made at a finite element square mesh whose lattice constant is  $a$ . Each square is then divided into two triangles so that the grid is in fact a triangulated mesh; by joining mesh points a 3D geometric structure of random peaks and valleys, each one is in a hexagon form that having six side surfaces, is constructed. Because now the surface is not as that defined by a simple condition  $z = 0$  for a smooth plane, we have to deal with the local surface plane in order to determine electron incidence/emission location. The reason to use a triangulated mesh is due to its advantage on easy judging the intersection of a velocity vector with a local triangulated plane when considering an electron incidence into the surface or emission from the surface (Ding & Li, 2005). A smooth plane surface can be constructed similarly by simply setting the vanishing amplitude. For the side surface of a line, it's also easy to construct a rough surface firstly on a virtual plane  $z = 0$  and then transform this plane with the mesh grid points into a declining side with the side wall angle  $\alpha$ . Fig. 10 shows the schematic diagram of a rough surface construction. Fig. 11 is an example of mesh building of a rough line, where each surface is roughness modulated.

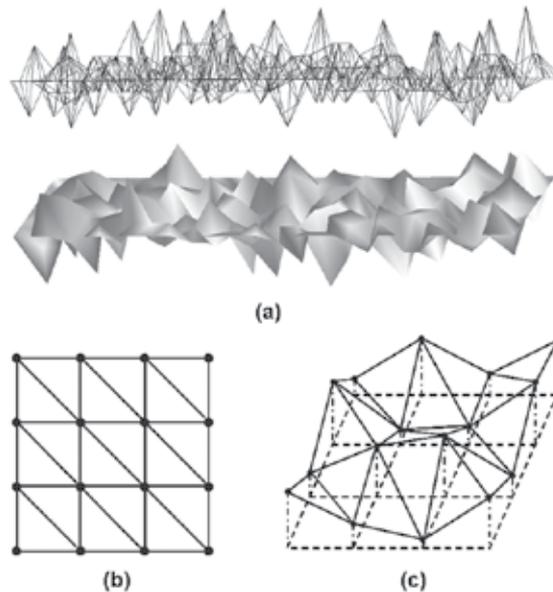


Fig. 10. (a) The schematic diagram of the roughness amplitude constructing by using triangulated meshes; (b) the triangulated mesh; (c) the local rough peaks constructed.

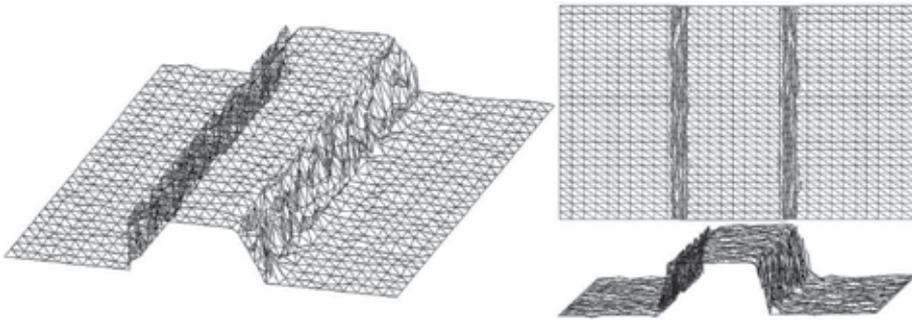


Fig. 11. Example rough surface structure of a trapezoidal line constructed with a triangulated mesh.

For the smooth structure, we can use two different constructive methods: first, the CSG modeling as well as a ray-tracing technique for SEM images for an inhomogeneous specimen with a complex geometric structure. Second, we can also construct the smooth geometry structure by using finite element triangle mesh while setting  $\sigma = 0$ .

Therefore, the geometric structure for a trapezoidal line with or without corner rounding, whose cross-sectional profile is shown in Fig. 12, is described by following parameters: the height  $H$ , the upside width  $W$ , the side wall angle  $\alpha$  and the corner radii  $r_U, r_D$ . The surface roughness can be parameterized by the amplitude  $3\sigma$  and the density described by the interval  $a$  of rough peaks. Using the parameter set,  $(H, W, \alpha, r_U, r_D, \sigma, a)$ , the linewidth can be easily described by the top-CD, (such as,  $W_T = W$ ), and the bottom-CD, (such as,  $W_B = W + 2 \tan \alpha$ ), as usual for the CD metrology of the gate lines.

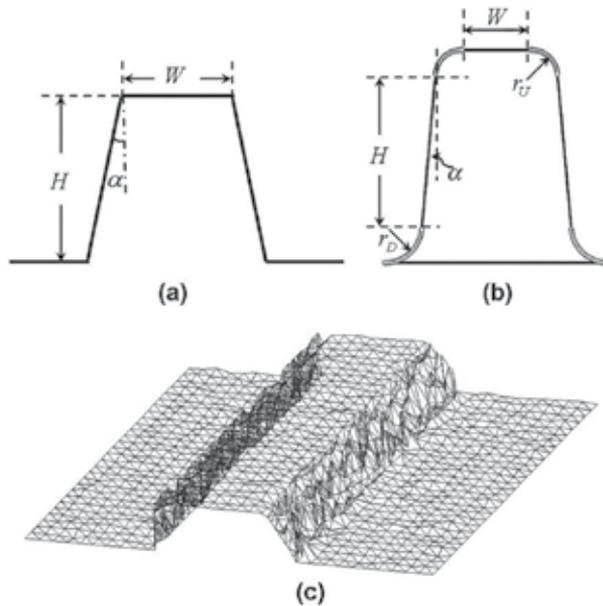


Fig. 12. The schematic diagram of specifying parameters of: (a) a trapezoidal line with smooth surface; (b) a trapezoidal line with corners and smooth surface; (c) a trapezoidal line with rough surface.

### 3.1.3 Computational conditions

The image pixels are set as  $225 \times 300$  or  $220 \times 500$  at an interval of 0.5-1.0 nm; for each pixel a number of  $10^3$  incident electron trajectories and several tens times of secondary electron trajectories are traced. To simplify the simulation the beam diameter is firstly assumed to be infinitely small; a more realistic image for finite probe size can be further obtained by a convolution procedure with a Gaussian probe width distribution. In the presented results we have not taken account of angular information of SE signals and also ignored any effect due to a detector and electric field in vacuum chamber, i.e. all the emitted electrons from the surface whose kinetic energy are smaller than 50 eV are taken as SE signals. A parallel computer is used to perform the calculation with a MPI program. The high efficiency of the program can be considered for the most complex case of surface roughness structures: The memory required is  $O(N)$ , i.e. proportional to the number of the rough surface peaks,  $N$ . The CPU time can be reduced from  $O(N)$  to  $O(N^{1/3})$  by introducing the spatial subdividing technique. For example, only a few minutes are necessary for a linescan and several hours for an image when using about 100 CPUs on a parallel computer.

### 3.2 Model validation

To verify the current MC simulation model, we have compared the simulated SE yield, the SE energy spectra and more directly the SEM line-scan and images with experimental results. The SE yield and energy spectra are fairly important to the SEM image simulation for it implicates the interaction mechanism of electrons with a solid. In our previous studies reasonable values of secondary yield and satisfactory energy distribution curves (Mao et al., 2008) as well as backscattering energy spectra (Ding et al., 2001; Ding et al., 2004b) have been successfully obtained, which has proved the reasonability of this MC model. For Si considered here, by using the full-Penn algorithm instead of single-pole approximation for the energy loss and considering more accurate SE excitation process, a good agreement had been obtained for absolute SE yield (Joy, 1995) and the SE spectrum (Joy et al., 2004) with experimental ones (Fig. 13).

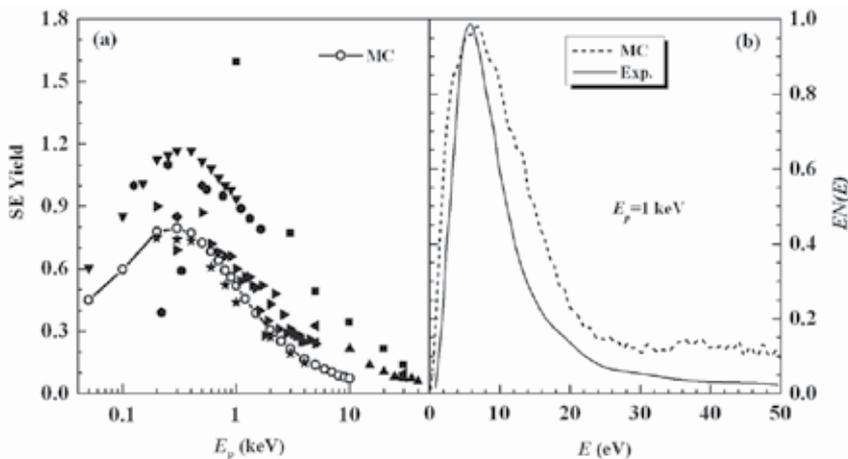


Fig. 13. Comparison on (a) the SE yield and (b) the SE energy spectra between experimental measurements and the calculations for Si. Experimental data are represented by solid symbols in (a) (Joy, 1995) and solid line in (b) (Joy et al., 2004).

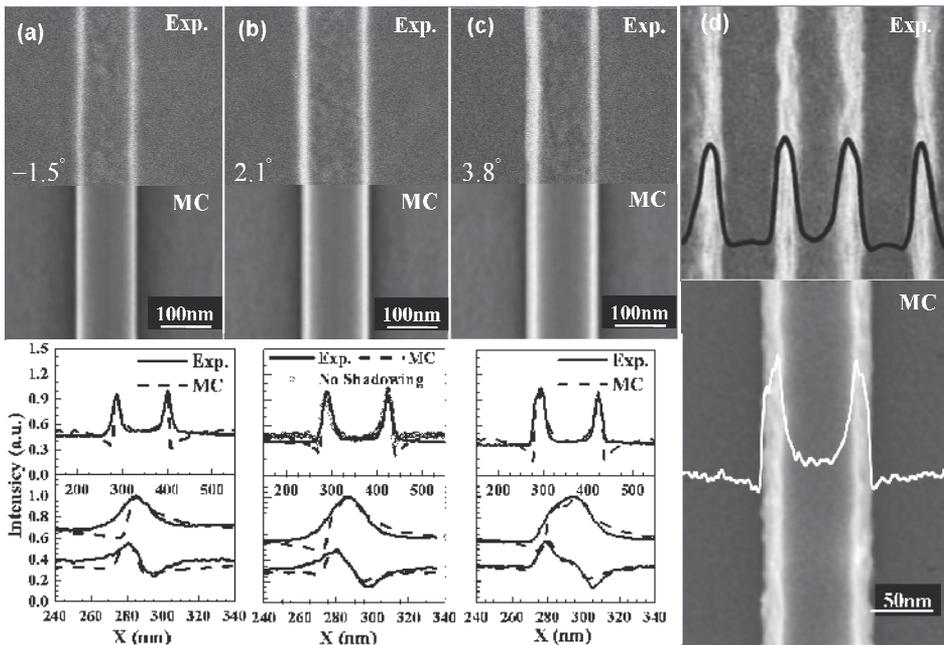


Fig. 14. A comparison on the SE line-scans and images between experimental results (Kawada et al., 2003; Shishido et al., 2002; Tanaka et al., 2003) and calculation. The input parameters in MC simulation are: (a)  $H = 240$  nm,  $W = 120$  nm,  $\alpha = -1.5^\circ$ ; (b)  $H = 240$  nm,  $W = 120$  nm,  $\alpha = 2.1^\circ$ ; (c)  $H = 240$  nm,  $W = 120$  nm,  $\alpha = 3.8^\circ$  for left side and  $\alpha = 2.1^\circ$  for right side; (d)  $H = 150$  nm,  $W = 65$  nm,  $\alpha = 5^\circ$ ;  $3\sigma = 6$  nm,  $a = 10$  nm for rough sides;  $3\sigma = 1.5$  nm,  $a = 2$  nm for rough top and substrate surfaces. The beam diameter for convolution is 8 nm for (a), (b) and (c), and 3 nm for (d). The primary energy of the electron beam is 0.8 keV for (a), (b) and (c), and 1 keV for (d).

Fig. 14 presents directly a comparison on SEM line-scans and images for smooth geometry lines between the simulation results for different parameters (sidewall angle and roughness) with experiments (Kawada et al., 2003; Shishido et al., 2002; Tanaka et al., 2003). The consistence found is reasonably well in different cases. The minor difference is that the simulated intensity drops at the edges of the line by the shadowing effect but in an experimental image this effect is vanished because of the existence of small external electric field. Fig. 14(b) indicates that, whether this shadowing effect is included or excluded dose not influence the linewidth. We can further deal with line edge roughness (LER) or line width roughness (LWR) as mentioned in Ref. (Li et al., 2008). It can be seen that the present MC simulated line scanning profile of a rough edge agrees qualitatively well with the experimental observation by choosing suitable parameters, indicating that the rough surface model introduced here is quite reasonable. One can thus estimate the actual roughness values of the observed system by comparing the simulated images with the experimental ones. The gate line segment distribution obeys Gaussian function with a standard deviation of line edge fluctuation (Bunday et al., 2004; Kim et al., 2004a; Xiong et al., 2004). Fig. 15 shows that, by a statistics made for the above example calculation of line scanning profile for a 65 nm gate line, a predicted Gaussian distribution of the linewidth has been indeed obtained. This fact also demonstrates that the rough surface model introduced here is very

reasonable. Here, the gate length is decided by only considering the two blooms of the line without introducing any experiential arithmetic because the usual arithmetic is questionable at nanoscale (Villarrubia et al., 2005a). It can be seen that the most probable value of the linewidth is a little bit ( $\sim 1$  nm) smaller than the actual geometry, which should be due to the inherent SEM image contrast formation. Hence, we can investigate the influence of each factor individually. Such a study will lead to an insight into building an applicable algorithm, perhaps the MBL method, for measuring linewidth.

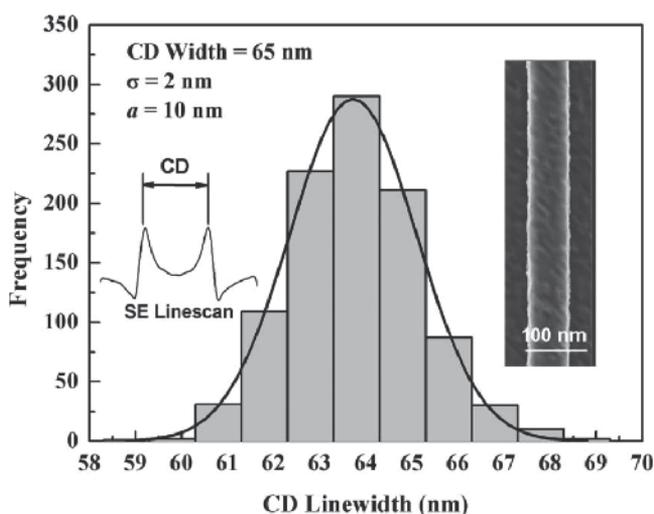


Fig. 15. The simulated gate line segment distribution of a rectangular Si line with the input geometry parameters:  $H = 150$  nm,  $W = 65$  nm,  $\alpha = 0^\circ$ ,  $3\sigma = 6$  nm and  $a = 10$  nm. The primary energy of the electron beam is 1 keV.

### 3.3 Influences of structural parameters to CDs

The precision of CD-SEM requiring optimum operation conditions with respect to stability, achievable resolution, signal-to-noise ratio and choice of primary energy (Frase et al., 2009). To clarify these conditions, in the following, we present the simulated line-scans of silicon lines and discuss the effect of each parameter one by one; they are electron beam parameters (energy, probe size and incidence angle) and geometry parameters (width, height, foot/corner rounding, sidewall angle and roughness).

The beam parameters are important factors to linewidth measurement. Probe energy, size, shape, current, incident angle and so forth can affect SE intensity and SEM contrast through different ways. Electron beam at low energies,  $E_0 \sim 300$ -3000 eV, is usually used in practice for it offers some important advantages to CD metrology by the following reasons (Cazaux, 2005): First, the low energy electron beam penetrates only into a thin surface layer in the specimen, thus reducing radiation damage and increasing the surface sensitivity of the imaging process. Second, the sharpness of images governed by topographic contrast can be increased for the extent of the edge effect decrease when the beam energy is low (Joy & Joy, 1996). Third, the SE yield is higher at low energies and, hence, the better signal-to-noise ratio at a given beam current and dwell time per pixel. To illustrate the influence of the beam energy to CDs in detail, the line-scans of a Si gate line at different beam energies are given in Fig. 16. The contrast increases with the energy. However, a little change on line-scan shape

has been found due to the change of interaction volume with primary energy. The CDs are found nearly the same for different beam energies. Thus, the beam energy factor influences the contrast and line-scan shape but none to CD.

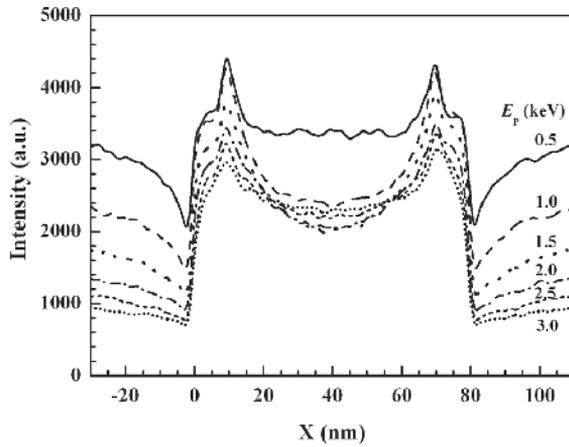


Fig. 16. The simulated SE line-scans of a Si line at different beam energies. The input parameters in MC simulation are:  $H = 150$  nm,  $W = 65$  nm,  $\alpha = 3.8^\circ$ , the beam size is 5 nm of FWHM.

The probe size and shape which describe the cross-sectional distribution of incident electrons can influence CDs dramatically. In SEM, the lateral resolution is also governed by the probe diameter (Bronsgest et al., 2008; Cazaux, 2005), which broadens the real dimension of the edges of a line. Probe size and current are interrelated (Bronsgest et al., 2008), thus, the resolution in practice is limited (e.g., about 3 nm in low energy region and less than 1 nm in much high energy region). To clarify the extent of influence of limited probe size to CDs, the line-scans of a Si gate line with different probe sizes are shown in Fig. 17.

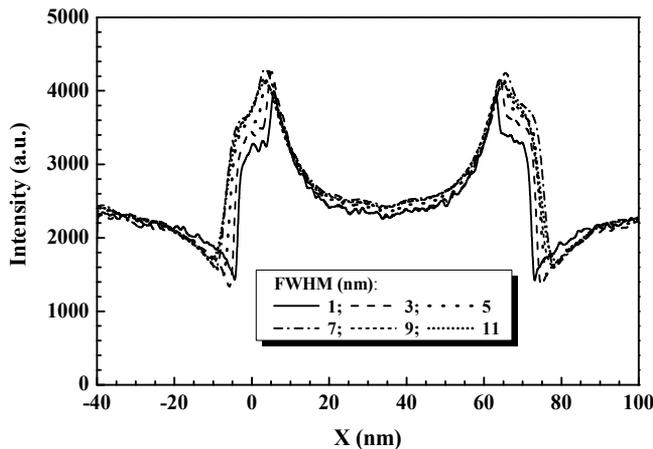


Fig. 17. The simulated CD-SEM line-scans of a single edge of a Si line with different beam sizes. The input parameters in MC simulation are:  $H = 150$  nm,  $W = 65$  nm,  $\alpha = 3.8^\circ$ . The primary energy of the electron beam is 1 keV.

The CDs increase gradually with the probe size. The bias of CDs caused by probe size can be estimated from this relation if the probe size is known. Here, the shape or the intensity profile of a probe is considered as a Gaussian distribution approximately for thermionic, Schottky and cold field emitters (Bronsgest et al., 2008). However, it has been pointed that some critical beam shape dependencies that are not correctly account for by the simple Gaussian model (Tanaka et al., 2005), and that the effect of electron incident angle is not negligible; the electron distribution far from focus is not consistent with a Gaussian model at the impact position of electrons on a line by the effect of aberrations of the beam shape. The requirement of the accuracy and stability of measurement in CD-SEM needs a better beam shape model (Tanaka et al., 2006).

The feature of a modeled line can be characterized by several pattern parameters. By considering the relationship of the simulated SE line-scan/image of modeled line with the experimental one, the pattern-dependent errors of linewidths could be removed. We have studied systematically different kinds of these pattern parameters, such as, width, height, foot/corner rounding, sidewall angle and roughness, and have discussed the influences to linewidth measurement quantitatively. In order to consider the contributions of different parameters, the beam conditions are kept the same in different cases as, the incident energy is 1 keV and the beam size is 5 nm of FWHM.

As shown in Fig. 18, when the width of the line is larger than 20 nm the SE line-scans look quite similar. However, when the width is smaller than the effective attenuation length the signal intensity will increase because of the edge effect. When the sidewall angle is not equal to zero, the model linewidth increases with the height; the line-scan edge shape and the distance between two blooms or two valleys can change dramatically. However, the height of the line does not influence the linewidth directly (Fig. 19).

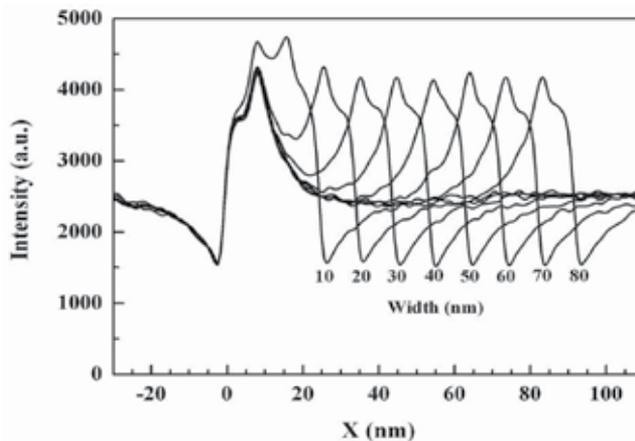


Fig. 18. The simulated CD-SEM line-scans for different widths of a Si line. The input parameters in MC simulation are:  $H = 100$  nm,  $\alpha = 5^\circ$ . The primary energy of the electron beam is 1 keV, the beam size is 5 nm of FWHM.

The reason for the negligible effect of width and the height on linewidth can be explained as follows (Reimer, 1998): the SE emission intensity at an edge obeys inverse cosine law as,  $1/\cos\theta$ , where  $\theta$  is a sidewall angle. Thus the CDs do not change with the width and height because the SE yield is the same on the edge surface where  $\theta$  is a constant.

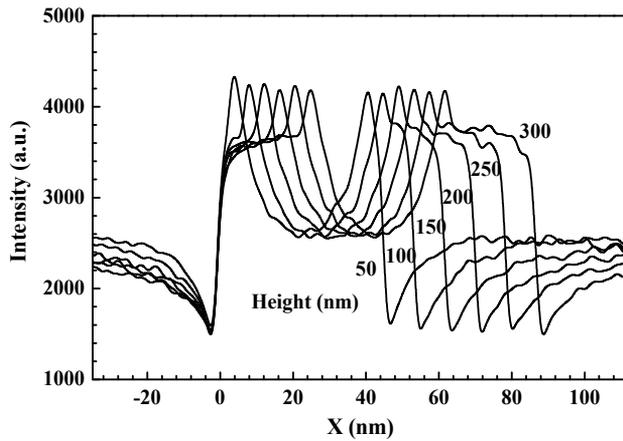


Fig. 19. The simulated CD-SEM line-scans of a single edge for different heights of a Si line. The input parameters in MC simulation are:  $W = 40 \text{ nm}$ ,  $\alpha = 5^\circ$ . The primary energy of the electron beam is 1 keV, the beam size is 5 nm of FWHM.

The line-scan shape at an edge is very sensitive to the sidewall angle. Thus, the sidewall angle is a critical factor to the linewidth measurement. To illustrate the dependence of linewidth on sidewall angle, the line-scans and also the bias between the simulation and model are shown in Fig. 20 where the sidewall angle is varied from  $0^\circ$  to  $9^\circ$ . The linewidth increases slightly with the sidewall angle increasing. The tendency is the same as the experimental results (Tanaka et al., 2003).

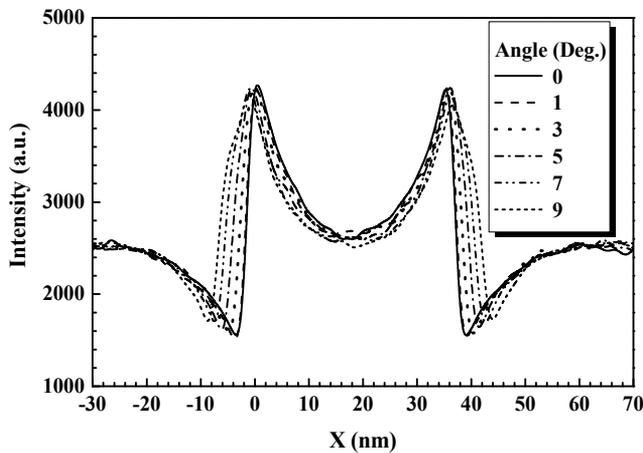


Fig. 20. The simulated CD-SEM line-scans for with different sidewall angles of a Si line. The sidewall angle index is defined in figure 9. The input parameters in MC simulation are:  $H = 40 \text{ nm}$ ,  $W = 40 \text{ nm}$ . The primary energy of the electron beam is 1 keV, the beam size is 5 nm of FWHM.

More generally, the line-scan shape is also sensitive to other angular features, such as, the footing and the corner angle indexes. The rounding feature of a line is usually represented quantitatively by the sidewall angle, the footing and the corner indexes (as shown in Fig. 21)

in practice (Morokuma et al., 2004). These two indexes can be defined based on the first derivative of the measured signal level, that is, the distances between peaks and outer zeroes of the first derivative are calculated as the feature indexes. As can be found in Figs. 22 and 23, both of the footing index and corner index are increasing with the bottom and top radius. The difference between them is mainly because a part of influence of sidewall angle index has been included in the corner index. The sidewall angle index and the corner index have the similar effect on the form of line-scans.

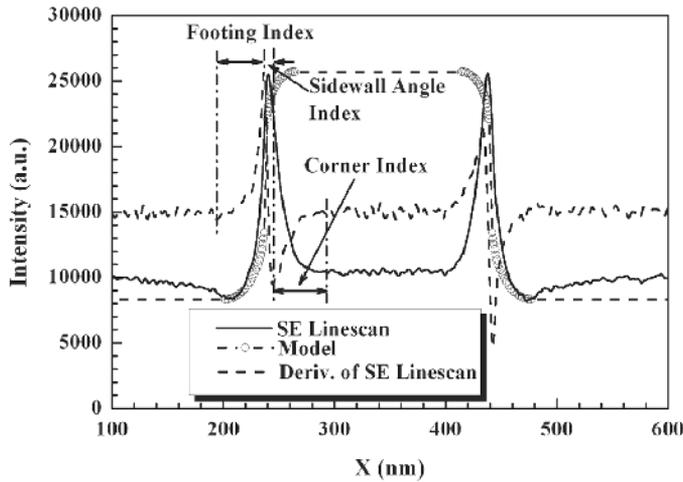


Fig. 21. The sketch for the definition of footing and the corner indexes of the linescan for a trapezoidal line with corner rounding as shown in Fig. 12(b).

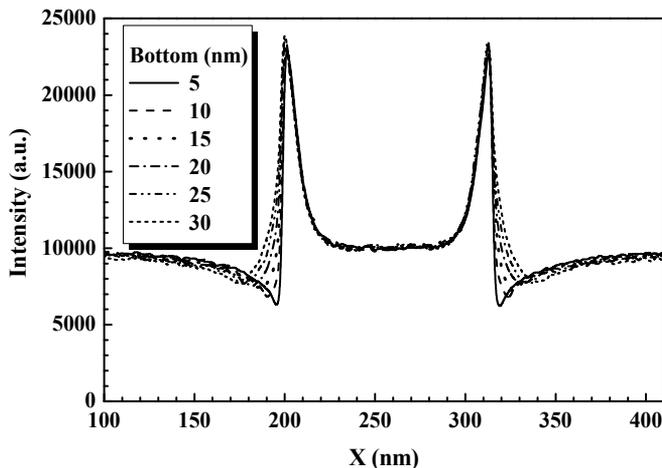


Fig. 22. The simulated CD-SEM line-scans of the lines with different bottom corner radius of a Si line. The input parameters in MC simulation are:  $H = 100$  nm,  $W = 40$  nm,  $\alpha = 0^\circ$ ,  $r_U = 10$  nm. The primary energy of the electron beam is 1 keV, the beam size is 5 nm of FWHM.

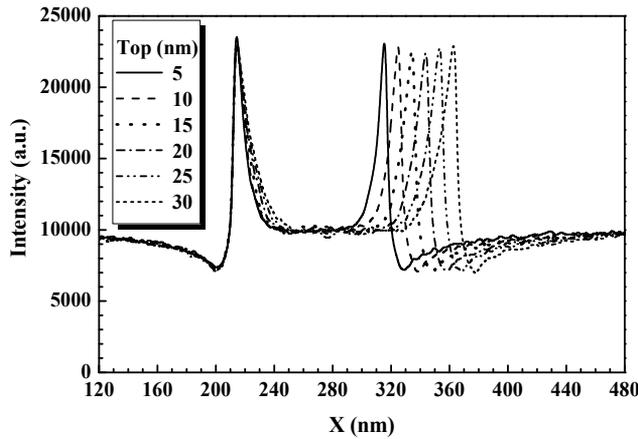


Fig. 23. The simulated CD-SEM line-scans of the lines for different top corner radius of a Si line. The input parameters in MC simulation are:  $H = 100$  nm,  $W = 40$  nm,  $\alpha = 0^\circ$ ,  $r_D = 15$  nm. The primary energy of the electron beam is 1 keV, the beam size is 5 nm of FWHM.

The LER and LWR have become a critical issue when the CDs for semiconductor devices shrink into a few tens of nanometers because it can degrade resolution and linewidth accuracy and cause fluctuation of transistor performance. A lot of experimental and theoretical researches have been done to confront this problem (Braun, 2005; Foucher et al., 2006).

In Figs. 24(a)-(e) and Figs. 25(a)-(e), the calculated CD-SEM images of 40 nm width lines are shown for different roughness amplitudes and densities, respectively. When the amplitude  $3\sigma$  increases from zero to 9 nm the line edge roughness of SEM images becomes obviously and the bias of the linewidth is also increased. However, the line edge roughness changes only slightly when the density decreases from  $1/2$  to  $1/9$  nm<sup>-1</sup>. The bias due to LER/LWR can be up to more than 10% for tens-nanometer linewidth. Indeed, it can become one of the main contributions to the bias of CDs in the future (ITRS, 2007).

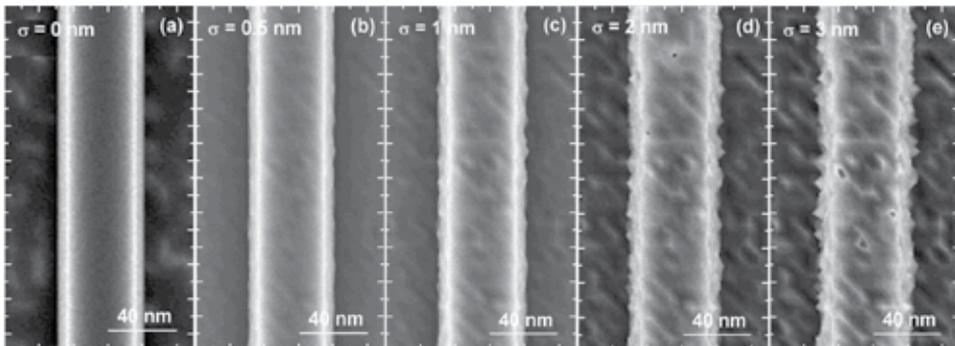


Fig. 24. The simulated CD-SEM images of the lines with different roughness amplitudes. The input parameters are:  $H = 100$  nm,  $W = 40$  nm,  $\alpha = 3^\circ$  and  $a = 6$  nm. (a)-(e): the  $3\sigma$  of roughness amplitude changes from 0 to 9 nm. The primary energy of the electron beam is 1 keV.

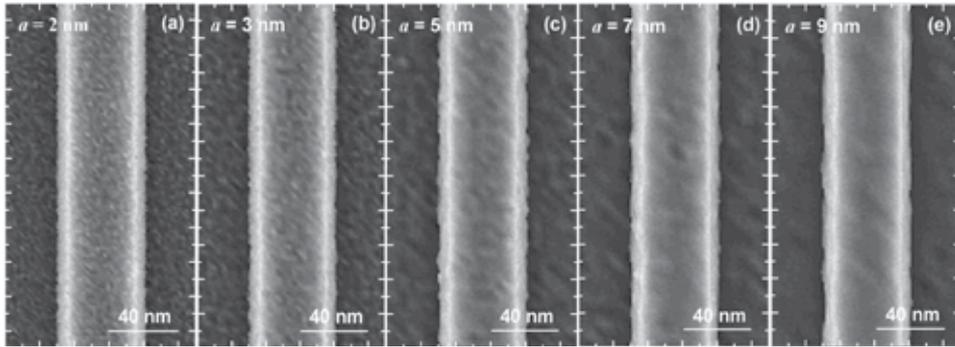


Fig. 25. The simulated CD-SEM images of the lines with different roughness densities. The input parameters are:  $H = 100$  nm,  $W = 40$  nm,  $\alpha = 3^\circ$  and  $3\sigma = 3$  nm. (a)-(e): the roughness interval changes from 2 to 9 nm. The primary energy of the electron beam is 1 keV.

The electrical charging phenomena play another major role in determination of CDs in CD-SEM for insulating specimens (such as, resist and  $\text{SiO}_2$ ) studied in semiconductor industry. Charging effects can change SE yield, contrast or other properties by its influence on the electron transport process and the surface potential. It is a dynamic process in the imaging by SEM, which can cause a dynamic component of CD measurement error (Babin et al., 2008a; Babin et al., 2008b; Cazaux, 2006). A lot of theoretical works have pointed out the role of charging effects in SEM (Babin et al., 2008a; Babin et al., 2008b; Cazaux, 1999; Cazaux, 2004; Renoud et al., 2004; Rau et al., 2008) and it has been considered in several MC models (Babin et al., 2008a; Babin et al., 2008b; Frase et al., 2009). It is found that a contrast reversal when beam voltage was varied; thus the charging effect can further influence the accuracy of CDs. The quantitative study of the influence of charging effects on CDs is urgent for the future work.

There are also other factors that may influence the CDs, such as, electron beam incidence angle and focus, material properties, SE signal detection and others (noise and etc.) (Babin et al., 2008a). Indeed, these factors can be ignored in most cases. For example, normal incidence beam are most used in CD-SEM, the changes of incidence angle are just used in some special cases or intentionally (Morokuma et al., 2004; Tanaka et al., 2003). Also, material properties or SE signal detection would change the SE yields then the contrast of images accordingly. Noise in a real CD-SEM image could produce measurement errors that have both random and nonrandom components (Xiong et al., 2004). The sensitivity of a roughness measurement to noise depends on both the choice of edge detection algorithm and the quality of the focus. Villarrubia et al. (Villarrubia et al., 2005a) have pointed out that measurements are less sensitive to noise when a model-based algorithm is used. In fact, the effect of noise can be studied by adding appropriate random noise to the simulated line-scan or more simply by changing number of incident electrons for the simulation because the noise and the roughness are uncorrelated.

As discussed above, the beam size, geometry of feature related to the angle indexes (sidewall angle, footing and corner angles), roughness and charging effects are the dominating factors to the CDs. Therefore, the accurate algorithm should be constructed by considering these factors theoretically. Furthermore, the errors attributed to different factors should also be distinguished quantitatively for different influences may cause the similar

line-scan feature; for example, the sidewall angles and height may cause quite similar feature when appropriate parameters are selected as shown in Fig. 26. A MBL algorithm can be constructed but it needs great amount of calculation corresponding to all possible values of parameters. Because the relationship of CDs with different factors is smooth and monotonous, we may fit the line-scan curves to construct a library of the relation of CDs with different factors.

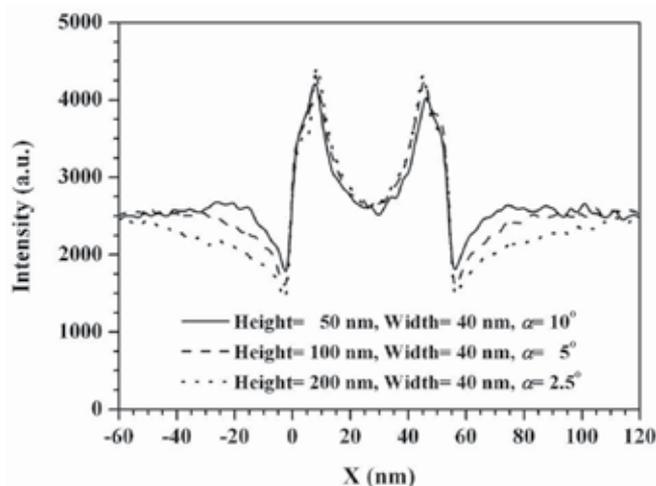


Fig. 26. A comparison of line-scans of lines with different side-wall angles and different heights. The primary energy of the electron beam is 1 keV, the beam size is 5 nm of FWHM.

#### 4. Simulation study of scanning Auger electron images

As a useful surface analysis tool, SAM has been used for elemental mapping of material surface. But, the quantitative mapping is quite difficult because some artifact signals (Prutton et al., 1995) can be produced in a complex process of electron beam interaction with a sample in addition to its disadvantages in the low signal-to-noise ratio, low spatial resolution and low energy resolution.

To comprehend the problems, many efforts (Cowley & Liu, 1993; El Gomati et al., 1978; El Gomati et al., 1979; Hembree et al., 1991; Hembree & Venables, 1992; Ito et al., 1996; Janssen & Venables, 1978; Liu & Cowley, 1993; Liu et al., 1993; Tuppen & Davies, 1985) have been done theoretically by using a MC simulation method and experimentally since ten years ago. Experiments were aimed to improve the spatial resolution for surface step and small particles deposited on substrates under different beam conditions (Cowley & Liu, 1993; Hembree et al., 1991; Hembree & Venables, 1992; Ito et al., 1996; Janssen & Venables, 1978; Liu & Cowley, 1993; Liu et al., 1993). The edge effect observed on a rough edge, which is caused by primary electrons when passing through the side surface of a step/particle and then hitting the surrounding substrate surface to generate extra Auger signals of the substrate elements, and the backscattering effect, which is resulted by primary electrons that backscatter to the sample surface to generate extra AEs of the surface elements, are mainly discussed for systems of thin film step on substrates by simulations of Auger line scans using MC methods (El Gomati et al., 1978; El Gomati et al., 1979; Tuppen & Davies, 1985). Recently, attentions (Jablonski & Powell, 2005; Powell, 2004) have been paid to the effect of

backscattered electrons on the analysis area in SAM for improving the precision in quantitative analyses. It has been also reported (Wight & Powell, 2006) a use of the extended logistic function for fitting AE and SE line-scans in order to provide a measure of interface width. A quantitative description of surface roughness effects on Auger peak-line profiles for pure and alloyed specimens was done in an experimental way (Agterveld et al., 1999). However, the important mechanism of Auger signal contrast has not yet been discussed in detail. Furthermore, the interested specimens have been shifted from micrometer structure to nanometer structure (Childs et al., 1996) now with the instrumental improvement (Jacka, 2001; Prutton, 2000; Venables & Liu, 2005); it thus requires explanations for many effects appeared in the nano-systems.

In this respect the image simulation can help us to comprehend the contrast formation mechanism and, therefore, is expected to play an important role for predicating the artifact and improving precision of elemental mapping by SAM. For this purpose we employ in the present work a MC electron trajectory simulation method, which can accurately describe the scattering and transport processes of incident electrons and of signal electrons beneath the sample surface (Ding & Shimizu, 1996; Shimizu & ding, 1992). Our previous comparisons (Ding et al., 2001; Ding et al., 2004a; Ding et al., 2004b) made on the energy distribution and yields of BSEs and of SEs have resulted very well agreement between MC simulation results and experimental measurements and, thus, confirmed that this MC physical model of electron scattering is quite reasonable. Though MC simulation technique has been widely used in studies of SEM and X-ray microanalysis (Gauvin et al., 1995; Yan et al., 1998) for simple geometrical specimens or even more complex geometric structures (Liu, 2000; Ly et al., 1995; Postek et al., 2002) to obtain high quality point analysis, line-scans and two dimensional images, however, MC simulation of SAM images is still very limited partly due to the difficulty to model a multi-elemental system with complex geometric structures.

The topic in this section then aims to extend the MC simulation of SEM images for complex sample geometries (as in Sec. 3) to that of SAM images. The physical model of electron scattering and SE generation is mainly based on that of Ding & Shimizu (Ding & Shimizu, 1996), and here we need to consider additionally the AE generation process. The geometric structure model of specimen by Li & Ding (Ding & Li, 2005; Li & Ding, 2005), which combines the CSG modeling and a ray-tracing technique, is used to treat an inhomogeneous specimen with a complex geometric structure (as in Sec. 2); each basic object for structure constructing can be chemically vacancy, element, alloy or compound. The size of a sample considered can be very small, such as in the order of nm, which is comparable to or even less than the electron scattering mean free path. Hence, necessary correction to sampling of electron scattering step length due to the specimen boundary condition must be considered. These improvements make it a meaningful MC model for AE image simulation of complex structured specimen surface.

#### **4.1 Simulation model**

The present MC model of electron scattering is mainly based on our previous approach (Ding & Shimizu, 1996), i.e. with the uses of Mott's scattering cross-section (Mott, 1929) for electron elastic scattering and Penn's dielectric function approach (Penn, 1987) to electron inelastic scattering. The cascade secondary electron production is included. However, we also need to consider the ionization events for AE generation, for which the Casnati et al.'s cross-section (Casnati et al., 1982) for inner-shell ionization has been used. The main feature of this model is given in Sec. 2 and only the inner-shell ionization, the cascade SEs/AEs generation and the special boundary corrections are outlined below.

**4.1.1 Inner-shell ionization**

Several empirical expressions of inner-shell ionization cross-section are commonly used for MC simulation of individual ionization events; among them the Gryzinski's formula derived from a classical binary collision model (Grizinski, 1965a; Grizinski, 1965b; Grizinski, 1965c) had been more popularly used simply because of its simplicity in the expression and completeness for providing excitation function, total cross-section and stopping power equation that are necessary for a simulation based on continuous slowing down approximation. However, the Gryzinski's cross section does not satisfactorily describe the experimental data. By careful examining several frequently used cross-sections, it has been concluded (Powell, 1989; Seah & Gilmore, 1998) that the best formula of inner-most shells ionization cross-section is that of Casnati et al. (Casnati et al., 1982). Therefore, we adopt the following Casnati et al.'s cross-section in this work,

$$\sigma_U = \frac{a_0^2 FR^2 AB \ln U}{UE_U^2} \tag{53}$$

This semi-empirical formula is the most suitable for describing the experiment data for  $2 < U < 60$ , where  $U = E/E_U$  is the overvoltage ratio,  $R$  is the Rydberg energy and  $E_U$  is the binding energy of an inner-shell,  $A$ ,  $B$  and  $F$  are the parameters related to  $E$  and  $E_U$ ,

$$F = \left(\frac{2+J}{2+T}\right)\left(\frac{1+T}{1+J}\right)^2 \times \left[\frac{(J+T)(2+T)(1+J)^2}{T(2+T)(1+J)^2 + J(2+J)}\right]^{3/2}; \quad A = (E_U/R)^d;$$

$$B = 10.57 \exp\left[(-1.736/U) + (0.317/U^2)\right]; \quad J = E_U/m_e c^2; \tag{54}$$

$$d = -0.0318 + (0.3160/U) - (0.1135/U^2); \quad T = E/m_e c^2 = UJ.$$

Fig. 27. presents the M-shell ionization cross sections of Au by Gryzinski's and Casnati's formulas, respectively.

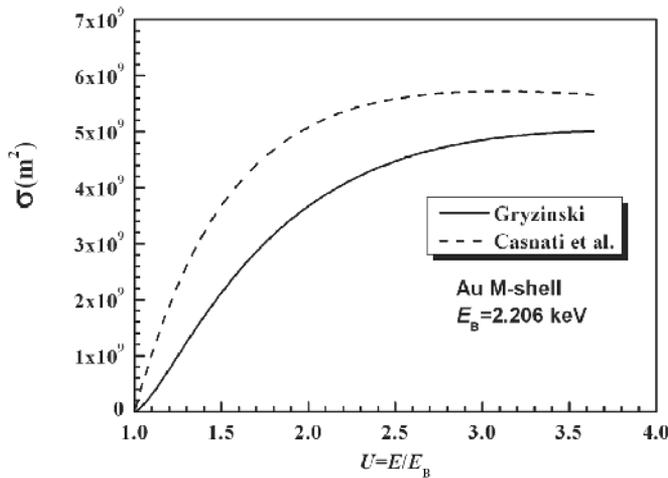


Fig. 27. The comparison of M-shell ionization cross sections of Au by Gryzinski's and Casnati's formula.

### 4.1.2 Cascade secondary electrons

In the present simulation model for discrete electron scattering events we have included SE generation because high energy SE may also slightly contribute to the inner-shell ionizations. Meanwhile the SE signals are essential to the simultaneous simulation of SEM image for comparison. We assume that each inelastic collision may produce a knock-on SE by transferring the loss energy  $\Delta E$ , to an inner-shell electron or a valence-conduction electron. The generated SE may undergo the similar inelastic collisions to cause a cascade SE production inside the sample. The dielectric function model mentioned above has been introduced to describe the energy loss; this cascade process is traced in the simulation until all the SEs either escape from the surface or come to rest within the sample. Only those emitted SEs whose energy is less than 50 eV are counted as true SE signals, otherwise they are BSE signals. The detailed description of this simulation process has been discussed elsewhere (Ding & Shimizu, 1996; Li & Ding, 2005; Yue et al., 2005).

### 4.1.3 Auger electrons

Simulation of AE trajectories and the associated scattering events can be treated in the similar way as for the incident electrons and the generated SEs. The difference only takes place for the initial condition of a trajectory start.

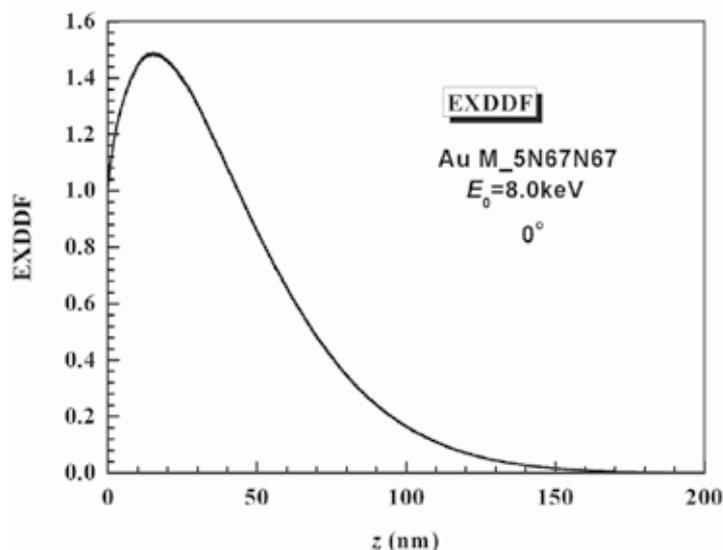


Fig. 28. The excitation depth distribution function (EXDDF) of  $\text{Au}(\text{M}_5\text{N}_{67}\text{N}_{67})$  for 8 keV at normal incidence.

In a MC simulation once an inelastic collision occurs, that is sampled by a random number with the cross-section ratio between inelastic one to the total one, there is a certain probability to be the inner-shell ionization event. Another random number determines the type of inelastic scattering channel according to the cross-section ratio of the inner-shell ionization to the total inelastic cross-section obtained in the dielectric function approach. Casnati et al.'s cross-section is then used in this process for determining the ionization probability; however, the energy loss is still determined by the differential cross-section in

the dielectric function approach when the fate of the inelastic channel selection is determined as inner-shell ionization. Certainly, here we need to consider which element and/or which shell is responsible for the ionization with the related inner-shell ionization cross-sections.

The signal intensity of AEs is quite weak in comparing with that of BSEs and SEs because of its rare probability. However, for a SAM image simulation we need only concern with relative intensity of Auger signals. Then we adopt a weighting method to increase the effective Auger signal intensity: in an ionization event we enable a certain amount (taken as 50 here) of Auger electrons, instead of only one, to be generated with the same characteristic energy at the same place of ionization location but their isotropic emission direction is randomly sampled. The energies, coordinates and directions of movement of these generated AEs are stored in the memory of a computer. After tracking all the trajectories of the primary electrons, the simulation for AEs and SEs will then be performed. Only those AEs emitted from the sample surface into the vacuum without losing much their characteristic energy, i.e.  $\Delta E < 1$  eV, will be registered as the concerned Auger signals. In Fig. 28, the excitation depth distribution function (EXDDF) of Au(M<sub>5</sub>N<sub>67</sub>N<sub>67</sub>) for 8 keV is given, which characterize the depth distribution of excited AE signals.

#### 4.1.4 Boundary corrections

A general simulation of AE image has to deal with chemically inhomogeneous specimen. For such a specimen the electron scattering mean free path is position dependent because the total scattering cross-section varies with spatial region. Then the conventional sampling for the step length should be modified when an electron passes through an interface separating different material zones. Especially when the feature size of the structure considered is under the order of submicron and is comparable with the magnitude of scattering mean free path, the modification to the sampling of step length should be particularly important. A ray-tracing technique (Ding & Li, 2005; Li & Ding, 2005; Yue et al., 2005) in the calculation procedure of electron steps has been employed here, which is suitable for a sample made of discrete elemental zones. The sampling step is now described as in Sec. 2 in detail.

Because the electron kinetic energy inside a material is referenced to the inner potential, it changes with the inner potential when an electron passes through the boundary separating different materials, as illustrated in Sec. 2 (Kotera et al., 1992).

#### 4.2 Results and discussion

MC simulations of line-scan profile and 2D image of AEs have been performed for several specimens. For calculation of a line-scan profile the trajectories of  $10^5$  primary electrons are simulated at each position of a line scanning across the specimen, and, for simulation of an image with  $200 \times 200$  pixels the trajectories of  $10^4$  primary electrons are tracked at each pixel. To simplify the factors influencing the images simulation the beam diameter is firstly assumed to be infinitely small; a more realistic image for finite probe size can be further obtained by a convolution with a Gaussian electron beam profile. In the presented results we have not taken account of angular information of signal electrons, i.e. all the emitted electrons from the surface are taken as respective signals according to their kinetic energy. We also ignore any effect due to a detector and electric field in vacuum chamber.

To verify the present MC simulation, the contrast investigation is firstly made by considering a particle/matrix system. Fig. 29(a) shows a quantitative comparison on the AE line-scans over Al particles on a Si substrate between the calculation and an experiment (Childs et al., 1996), and Fig. 29(b) shows a comparison for 2D SAM images of Al particles, treated as three semi-spheres of different diameters in calculation, placed on a Si substrate surface; the same condition as in the experiment (a 20 keV primary electron beam and the CMA analyzer) was considered in the calculation. By adjusting the parameters, i.e. particle diameter and beam diameter, in the simulation line-scan profile fit the experimental results very well with the estimated beam size of 20 nm. Furthermore, the contrast of the calculated Al(KLL)-AE image for a 20 nm beam spot size is also very close to that of the experimental image.

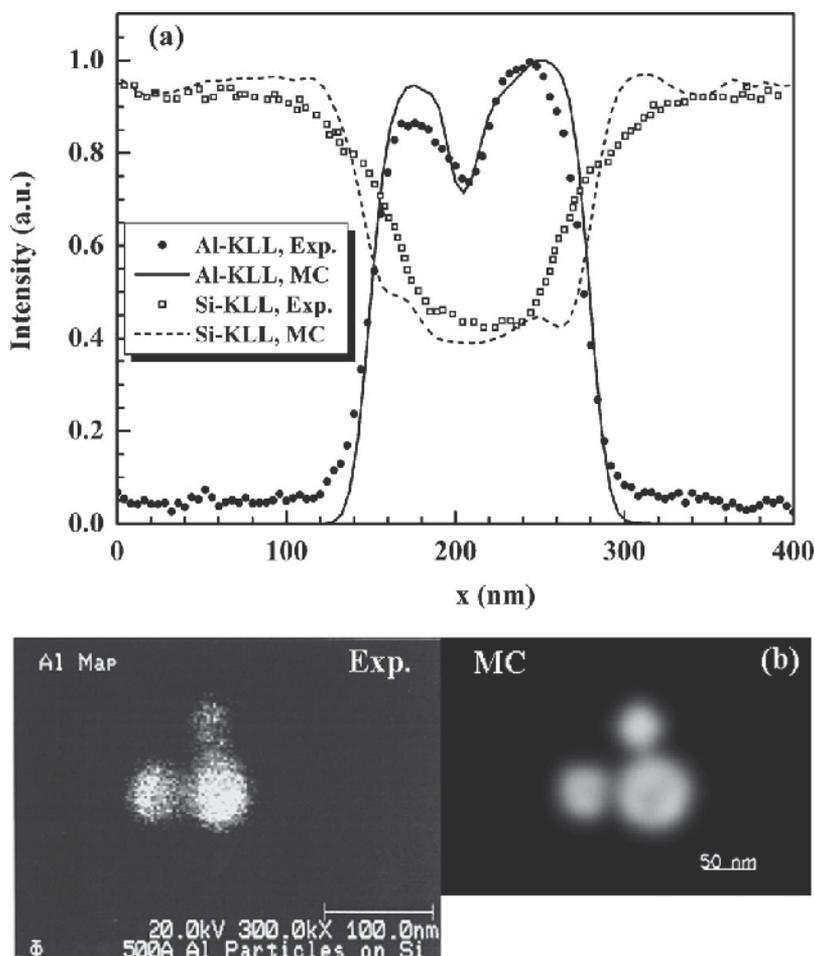
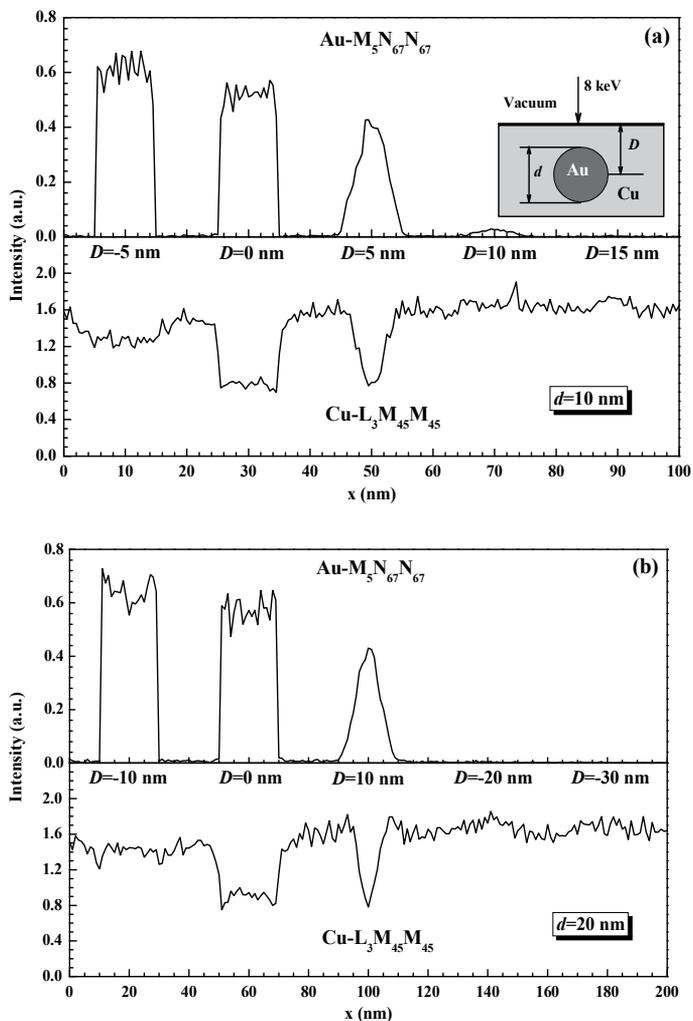


Fig. 29. Comparison on the Auger (a) line scans and (b) images of Al semi-spheres placed on a Si surface between the Monte Carlo simulation and an experiment (Childs et al., 1996). The primary electron beam of 20 keV is at normal incidence to the surface; the beam size is estimated to be 20 nm. The diameters of three semi-spheres are 56, 76 and 44 nm. The Al(KLL) - and Si(KLL) -Auger signals are measured with a CMA analyzer.

The simulations were then performed for different conditions by varying particle composition, size and location, the primary energy and the incident angle. For the contrast of such a system the topographic factor may also be a main source of contrast of an AE image. For quantitative surface chemical mapping, it is intended to reduce the artifacts by reducing topographical contrast and revealing otherwise hidden chemical contrast (Prutton et al., 1995). However, there are still no confirmed conclusions about the effect of each factor till now.

In Fig. 30 the calculated Auger line-scan profile is shown for an Au sphere located on or embedded in a Cu matrix. An obvious contrast change for both Au( $M_5N_{67}N_{67}$ ) and Cu( $L_3M_{45}M_{45}$ ) AE signals can be found when changing the depth of Au sphere as well as the sphere diameter. One can see that, for Au( $M_5N_{67}N_{67}$ ) Auger line-scans, the contrast due to topography retains obviously when a particle is placed on the surface or half embedded in the matrix. On the other hand, when an Au particle is entirely embedded in the matrix so that the topography factor is further weakened, the contrast varies quickly



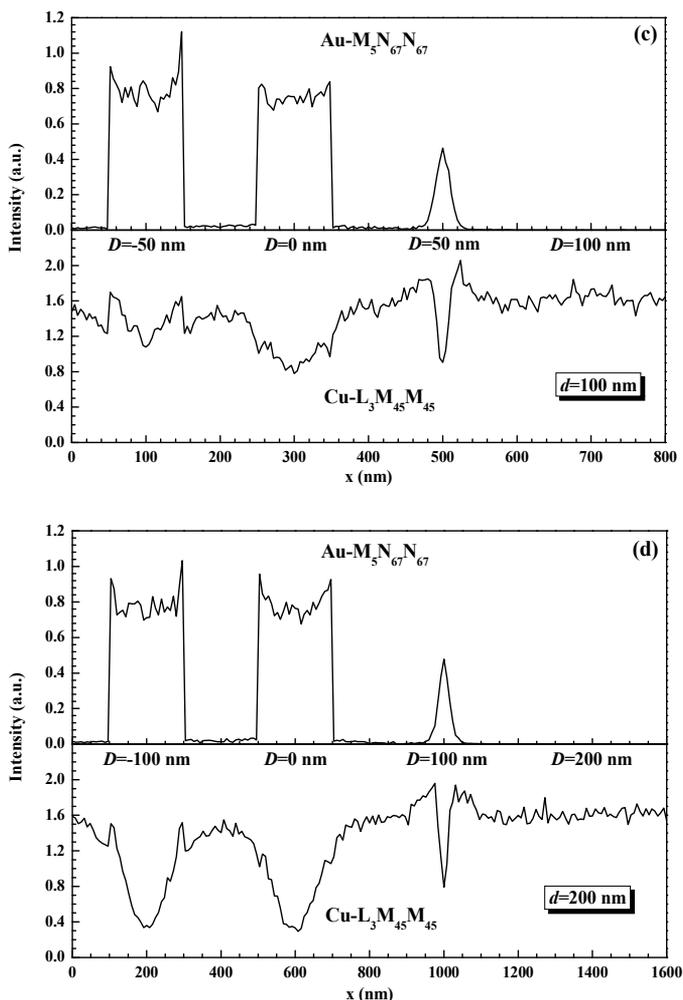


Fig. 30. Calculated  $\text{Au}(\text{M}_5\text{N}_{67}\text{N}_{67})$ - and  $\text{Cu}(\text{L}_3\text{M}_{45}\text{M}_{45})$ -Auger line-scans for an Au sphere located on a Cu surface or embedded in a Cu matrix.  $D$  is the depth of the particle (the negative value of  $D$  represents the case of a particle placed on the surface).  $d$  is the diameter of the particle: (a) 10 nm; (b) 20 nm; (c) 100 nm; (d) 200 nm. The primary electron beam of 8 keV is at normal incidence to the surface.

with the depth and wholly vanishes when the top of the particle reaches the signal effective depth, typically 0.3~3 nm. This is obviously due to the emission range limitation for AEs that excited by such a primary beam of 8 keV. Only those AEs generated from the outmost atomic layers of a solid surface can survive to be ejected and registered as AEs; this range estimated by IMFP depends on the AE energy only. The profile shape alters not only with the depth of a particle but also with the particle size. For large particles ( $d = 200$  nm) there is an edge contrast quite similar to that observed in SEM. The reason for this sharp increase of Auger signals at an edge is that AEs generated can have greater chance of emission from the side surface when an electron beam impacting on the edge. Therefore, topographical factor plays also an important role in the SAM contrast of large features. For

$\text{Cu}(\text{L}_3\text{M}_{45}\text{M}_{45})$  Auger line-scans, the contrast is expected to be opposite to that of  $\text{Au}(\text{M}_5\text{N}_{67}\text{N}_{67})$ . This is almost true except in the case that a small particle (e.g.  $d < 20$  nm at 8 keV) is placed on the surface (Fig. 30(a)): the dip of Cu-Auger electron current at the Au particle region is so weak that it is hard to observe the placed small Au particles from the variation of the substrate Auger signals. This is also a topographical effect for small features. For submicron Au particles (e.g.  $d > 100$  nm at 8 keV), the usual edge enhancement effect (caused by higher-energy SEs that passing through the side surface of a particle and hitting the substrate to generated extra AE signals of the substrate (Ito et al., 1996)) and shadowing/shading effect (shadowing of the AEs on their emission path by a particle leads to a decrease of the substrate signals at the edge (Shimizu & Everhart, 1978)) become obvious in Cu-Auger line-scans as observed experimentally (Ito et al., 1996). These effects due to the excitation and emission process of AEs have been well expounded (El Gomati et al., 1988; Shimizu & Everhart, 1978; Tuppen & Davies, 1985; Umbach & Brunger, 1989; Wells, 1974).

To separate the topographical factor from the chemical factor in contrast a SAM image, Fig. 31 illustrates the contrast purely due to topographical factor for the case of Cu particles placed on a Cu substrate. The calculated contrast is then quite close to that of a secondary electron SEM image (Yue et al., 2005). This fact shows clearly that the topographical factor plays also an important role in SAM image contrast when the particle size is larger than the attention depth of AEs, leading to the edge effect. Rough surface, which could lead to more AE signal emission by enlarging the effective surface region, consequently, enhances the intensity of AE signals and produces the topographical contrast.

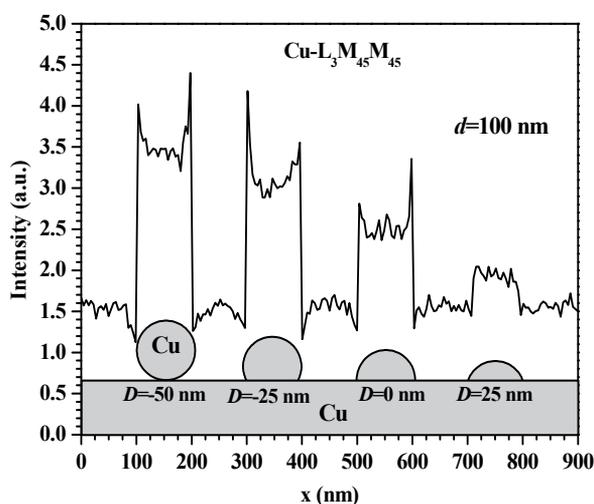


Fig. 31. Calculated  $\text{Cu}(\text{L}_3\text{M}_{45}\text{M}_{45})$ -Auger line-scan for Cu spheres located on a Cu surface or embedded in a Cu matrix. The diameter of spheres is 100 nm and the separation between them is 200 nm. The primary electron beam of 8 keV is at normal incidence to the surface.

The chemical factor in SAM image contrast is mostly related to the dependence of excitation and emission processes of AEs on atomic number of materials. The contrast is basically caused by the Auger excitation probability of particular atoms within a small region of electron beam impact area; when the interaction volume of electron beam is larger than the

feature size of a structure the creation and transportation of AEs in a nearby spatial region which is chemically different could become an important factor to influence the contrast. Therefore, chemical and topographical factors are both dominant to the contrast.

The contrast properties presented above can be observed more intuitively by a 2D image. Fig. 32 shows the calculated results for Au particles of 100 and 200 nm diameters at different depths in a Cu matrix for a primary energy of 8 keV. Au( $M_5N_{67}N_{67}$ ) and Cu( $L_3M_{45}M_{45}$ ) AE images are both obtained. In Au-Auger electron images (Figs. 32(a) and (c)), the contrast obviously becomes weaker as the particle locates deeper inside the matrix, and, it even entirely vanishes when the depth reaches about the particle diameter. When the particle is located on the surface or half-embedded in the matrix, a very clear and similar contrast can be observed. In substrate Cu-Auger electron images (Figs. 32(b) and (d)), a corresponding contrast is observed. The edge enhancement effect and the shadowing/shading effect can also be clearly seen in the 2D image. It obviously shows that the contrast of Cu-Auger electron signals for a particle placed on the surface is higher than that half-embed in the matrix. This is due to the fact that the topography factor becomes more crucial so that the edge enhancement effect becomes more obviously when the particle placed on the surface.

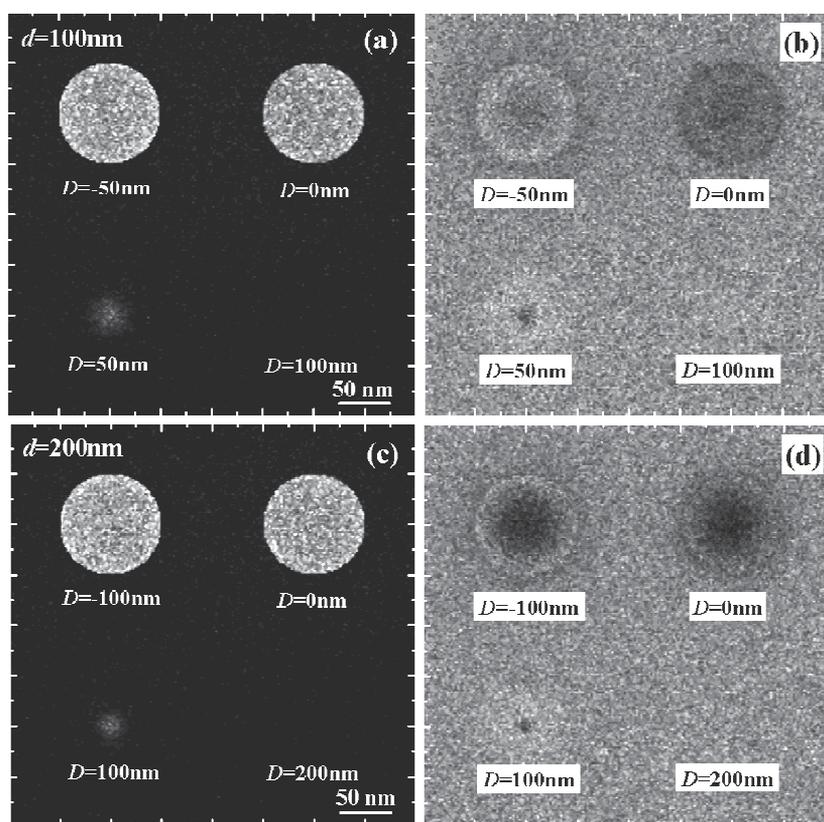


Fig. 32. Simulated SAM images for Au particles located at various depths in a Cu matrix: (a) Au( $M_5N_{67}N_{67}$ ),  $d = 100$  nm; (b) Cu( $L_3M_{45}M_{45}$ ),  $d = 100$  nm; (c) Au( $M_5N_{67}N_{67}$ ),  $d = 200$  nm; (d) Cu( $L_3M_{45}M_{45}$ ),  $d = 200$  nm. The primary electron beam of 8 keV is at normal incidence to the surface.

The contrast and the resolution of AE image are two main factors concerned by SAM. It has been shown that utilization of the low primary beam voltage of 3 kV has the advantage of reducing the edge effect in analyzing a  $0.7 \mu\text{m}$  TiN particle on a steel (Forsyth & Bean, 1994; Olson et al., 1993) and an Au bar ( $0.6 \mu\text{m}$  high and  $1.0 \mu\text{m}$  wide) on a Si substrate (Tuppen & Davies, 1985). While it has also been pointed out that (Ito et al., 1996) electrons with high energy will penetrate deep into the particle to reduce edge effect. The edge effect can dramatically degrade spatial resolution (El Gomati et al., 1988; Shimizu & Everhart, 1978; Tuppen & Davies, 1985) because AE from substrate can be detected even when the primary beam is impacted at the particle. At lower primary energies, it is more difficult for electrons to pass through the side surface of a particle and to hit the surrounding substrate surface for producing the edge effect. At higher primary energies, incident electrons can penetrate much deep into the particle to reduce the edge effect. Both of the extreme cases can improve the resolution of Auger images by reducing the edge effect.

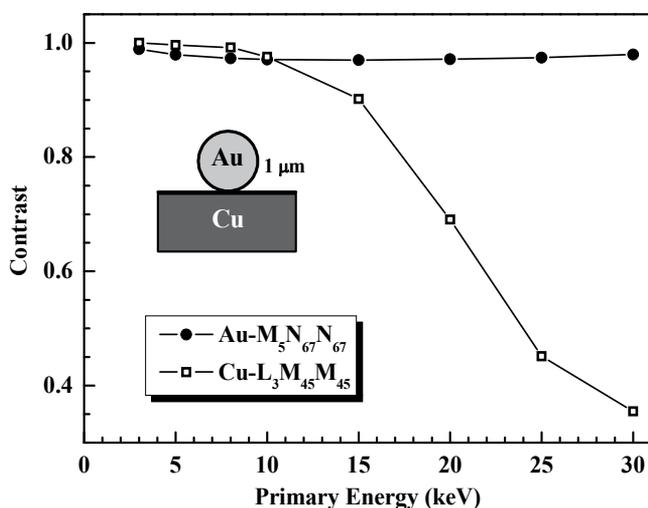


Fig. 33. Simulated Au( $M_5N_{67}N_{67}$ )- and Cu( $L_3M_{45}M_{45}$ )-Auger signal contrasts for an Au particle ( $d = 1 \mu\text{m}$ ) placed on a Cu surface as a function of energy of a primary electron beam at normal incidence to the surface.

To investigate the accurate relationship between the contrast of AE image and primary electron energy, we have performed simulation of Auger line-scan of an Au sphere of the diameter of  $1 \mu\text{m}$  placed on a Cu substrate surface at different primary energies and at a normal incidence condition (Fig. 33). The simulations of the line-scan profile have enabled the contrast to be estimated, which is defined as  $C = (I_{\text{max}} - I_{\text{min}}) / (I_{\text{max}} + I_{\text{min}})$ , where  $I$  denotes the AE intensity. It can be seen that with the increase of the primary electron energy (3-30 keV), the Au-Auger electron contrast firstly decreases a little bit to a minimum at about 15 keV corresponding to overvoltage ratio of  $U = 7 - 8$  and then increases. But the overall contrast in the energy range is nearly unit because the Au-Auger electrons can only be excited at the position of Au particle. However, the contrast for substrate Auger signals decreases consistently with increasing primary energy. This is due to the edge effect and backscattering effect that mentioned previously (El Gomati et al., 1988; Shimizu & Everhart, 1978; Tuppen & Davies, 1985). Both of these two effects can reduce the intensity difference

of the Cu-Auger electrons when a primary beam scans over a particle. The relations found here are useful to estimate the proper energy for the best contrast observation.

Fig. 34 shows the dependence of contrast on particle size. With increasing particle size the contrast of Au- $(M_5N_{67}N_{67})$  Auger image falls to be a constant but still nearly equals to unit, while the Cu- $(L_3M_{45}M_{45})$  contrast rises. This relation between contrast of Au signals and particle size can be easily understood from the mechanism of shading effect and the AE emission process. The contribution to the emitted Au-Auger signals by those electrons generated from deeper depth than IMFP can be neglected. Most of the emitted AEs are produced within a very short distance from the sample surface, typically 0.3~3 nm, which is enough small compared with the particle diameter concerned here. Therefore, the contrast of Au-Auger image does not change almost with particle size. For Cu-Auger electron image, the enhancement of the contrast with increase of particle size is mainly due to the reduction of substrate AE emission from the position underneath the particle. The topographical factor has dominated the Cu-Auger contrast: on the one hand, the edge effect is un conspicuous when the particle is very small (< 10 nm); on the other hand, for large particles (>400 nm) the contrast is a constant (~1 if the background signal is ignored) for the shading effect becomes obvious. In the middle range of the particle size (10~60 nm) where the contribution to the contrast by the edge effect exceeds that by the shading effect, a contrast reversion would appear. In the other words, for these particle sizes the detected intensity of substrate AEs when a primary beam is impacted at the center position of the particle can be even higher than that when a primary beam is impacted at the substrate surface. This effect is mainly due to that more substrate AEs can be excited by the edge effect, but less of these AEs emitted from the substrate would be shadowed by the particle.

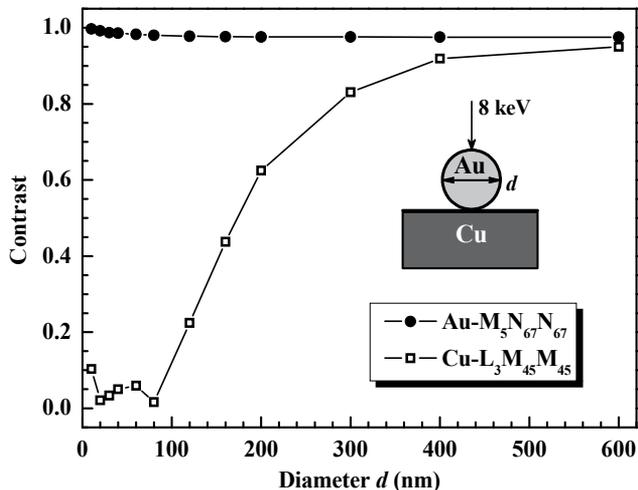


Fig. 34. Simulated Au( $M_5N_{67}N_{67}$ )- and Cu( $L_3M_{45}M_{45}$ )-Auger signal contrasts for an Au particle placed on a Cu surface as a function of particle size for a primary electron beam of 8 keV at normal incidence.

In order to investigate the contrast more comprehensively, we have also performed a calculation for the oblique incidence condition at an incident angle of  $45^\circ$ . The results for line-scans of an Au particle of diameter 100 nm located at different depths in a Cu matrix are

illustrated in Fig. 35. The following effects can be observed: Firstly, the Auger intensity peaks and valleys have a position shift. This is due to that the line-scan position does not correspond linearly to the true impact location at structures protruding from a plane surface when they are illuminated by an oblique electron beam. Secondly, the shapes of line-scan profiles are different for different depths because of an obstructing effect of Au particles. This difference in contrast may be helpful in judging the in-depth position of particles for such a system. Furthermore, the shading effect and the edge effect become more obvious in the forward side and the backward side, respectively. This is mainly because of that the oblique incidence condition can decrease the backscattering effect in the forward side facing electron beam but increase the effect in the backward side.

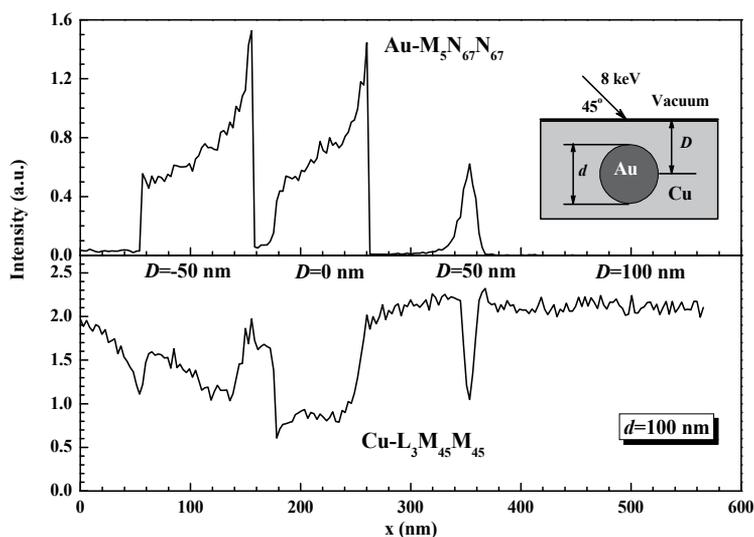


Fig. 35. Calculated  $\text{Au}(\text{M}_5\text{N}_{67}\text{N}_{67})$ - and  $\text{Cu}(\text{L}_3\text{M}_{45}\text{M}_{45})$ -Auger line scans for an Au particle ( $d = 100$  nm) placed on a Cu surface and for a primary beam of 8 keV at an incident angle of  $45^\circ$ .

Besides the above effects, several other effects are studied for the systems under the special conditions. These effects are the contrast enhancement for different elemental particles that wholly embedded in a matrix, the artifact contrast due to nearby geometries containing different elements, and the variation of substrate AE intensity by the chemical composition of tiny particles. The mechanism for all of these effects can be explained by the electron transport and scattering processes under different special geometry configurations.

In Figs. 30 and 32, an enhancement of  $\text{Cu}(\text{L}_3\text{M}_{45}\text{M}_{45})$ - AE signal in a line-scan or an image has been shown when an Au particle is wholly embedded in a Cu matrix. The effect is contrast to one's expectation and should be mainly due to the difference on the backscattering probability for the incident electrons between elements Au and Cu. If a nanoparticle of heavier element is embedded in a matrix of lighter element, the intensity of the matrix Auger signals would increase because more BSEs from the embedded particle would transport to the nearby matrix region where they can excite more matrix AEs. This effect is useful for detection of particles underneath a surface by increasing the primary electron energy.

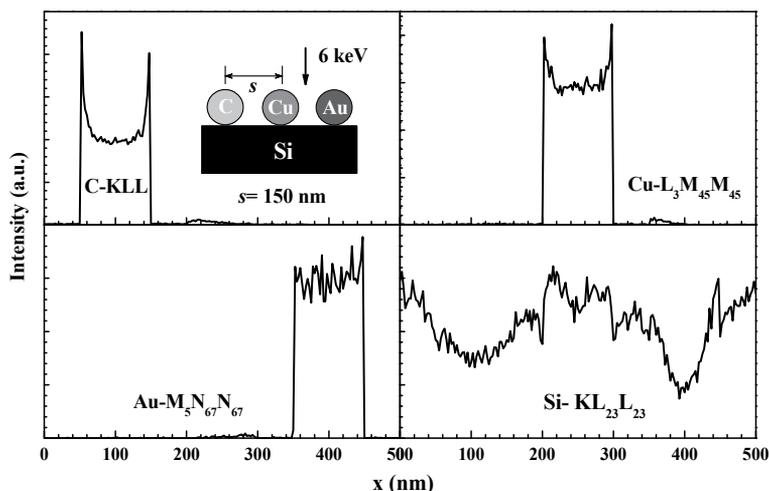


Fig. 36. Calculated line-scan of Auger electrons for a specimen made of C-, Cu- and Au-spheres placed on a Si surface and for a primary beam of 6 keV at normal incidence: (a) C(KLL); (b) Cu(L<sub>3</sub>M<sub>45</sub>M<sub>45</sub>); (c) Au(M<sub>5</sub>N<sub>67</sub>N<sub>67</sub>); (d) Si(KL<sub>23</sub>L<sub>23</sub>). The diameter of spheres is 100 nm and the separation between them is 150 nm.

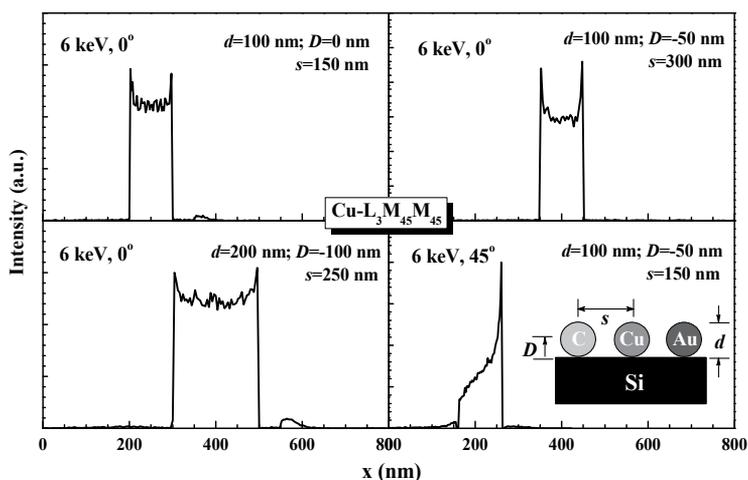


Fig. 37. Calculated Cu(L<sub>3</sub>M<sub>45</sub>M<sub>45</sub>)-Auger line-scans for a specimen made of C-, Cu- and Au-spheres placed on a Si surface or half-embedded in a Si matrix and for a primary beam of 6 keV.  $D$  is the depth of a particle (the negative value of  $D$  represents the case of a particle placed on the surface).  $d$  is the diameter of a particle and  $s$  is the separation between particles.

Fig. 36 illustrates line-scan profiles, for several different elemental particles closely placed on a surface of another elemental substrate, of corresponding elemental Auger signals. An important effect causing an artifact contrast of Auger images has been found, e.g. C(KLL) signals appear in the Cu-particle region and Cu(L<sub>3</sub>M<sub>45</sub>M<sub>45</sub>) signals in the Au-particle region. To explain the origin of artifact signal let us now, for example, consider Cu(L<sub>3</sub>M<sub>45</sub>M<sub>45</sub>) signals: when an electron beam scans over a heavy elemental Au-particle. Some of incident electrons can be emitted from Au-sphere after suffering multiple elastic scatterings to deflect

largely from their incident direction, and thus reentered into the nearby Cu particle; these scattered electrons can then have enough energy to produce Cu-Auger electrons inside the Cu particle, and those emitted Cu-Auger electrons are recorded as artifact signals at the scanning location of Au-particle. However, when an electron beam scans over a light elemental C-particle, less incident electrons can be scattered out of the particle with large scattering angle and, thus, produce negligible artifact signal at the scanning location of C-particle region. For substrate Si(KL<sub>23</sub>L<sub>23</sub>) signal electrons, they are less produced in the region below Au-particle and many of them are hardly to be emitted from the region due to blocking of the particle. It can be seen that many factors may influence the artifact contrast, such as, the particle elements and respective Auger signal, the separation between particles, the particle size and depth, the incident energy and angle of incidence. Different conditions have thus been considered in simulation. Figs. 36(b) and 37(b) clearly show that the artifact contrast decreases dramatically when the separation between two particles change from 150 to 300 nm. Fig. 37(a) indicates that the artifact contrast still presents when the particle is half-embedded in the matrix. The artifact signal intensity increases with particle size in Fig. 37(c). For an oblique incidence, the artifact signal can be increased in the region of forward side but decreases in backward side, as shown in Fig. 37(d). Furthermore, it has been shown that the artifact C-Auger signals are more obvious at the position of Cu-particle but the artifact Cu-Auger signals are nearly vanished at the position of C-particle; the difference varies with the differences between atomic numbers of two relative elements of particles.

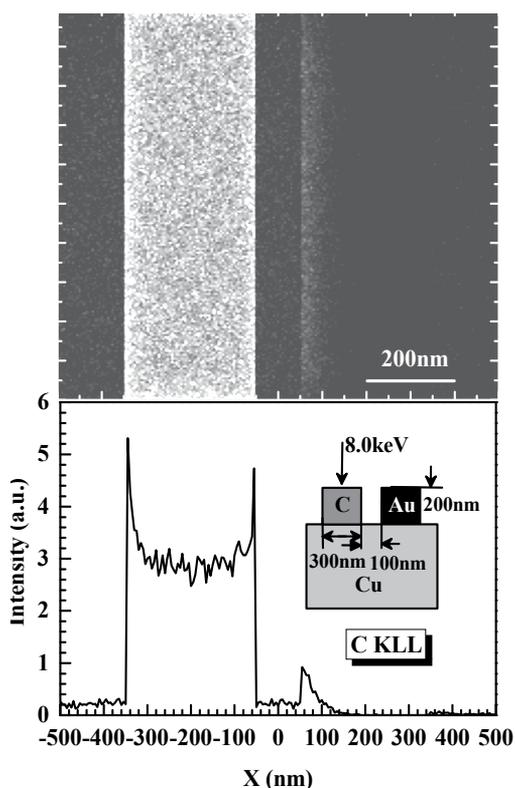


Fig. 38. Simulated C(KLL) -Auger line-scan and image for a specimen made of C- and Au-cuboids on a Si surface and for a primary beam of 8 keV at normal incidence.

Hence, as discussed above, the geometrical structure and elemental composition of specimen, the electron beam condition and Auger excitation shell are those factors to affect the artifact contrast of such a system. This artifact phenomenon is basically due to the backscattering of electrons. Obviously, it is much easier for electrons to escape from the sharp edges and to produce a stronger artifact contrast. Fig. 38 shows a simulated line-scan profile and a 2D image of C(KLL) signal for a system made of C- and Au-cuboids with small separation on a Si substrate. The result clearly shows that a sharp edge can strengthen the artifact contrast, especially for a nanometer structure.

In Fig. 36(d), another effect that the substrate Si-Auger electron intensity varies with the chemical composition of tiny particles has been predicted. The effect can be more obviously displayed in a 2D image as shown by Fig. 39. The contrast is due to many factors that may influence the transport property of signal electrons, such as, density, atomic number and thus stopping power and scattering cross-sections etc. Fig. 39 also shows the simulated SEM images of SEs and of BSEs in order to compare with the substrate SAM image. The signal intensity in a backscattered electron image increases only with atomic number through electron elastic scattering cross section while in a secondary electron image relates mainly to the stopping power of electron inelastic scattering. There is an additional difference between the substrate SAM- and SEM-signals: the substrate SAM signals are produced only inside the substrate and are blocked by the particles during their emission; therefore, the signal can only appear for small particles and it will vanish when the particle size is all greater than the attenuation length of substrate AEs, as shown by Fig. 40. Obviously, this calculation may help to understand the contrast formation mechanism of a SAM image of such nanometer structures.

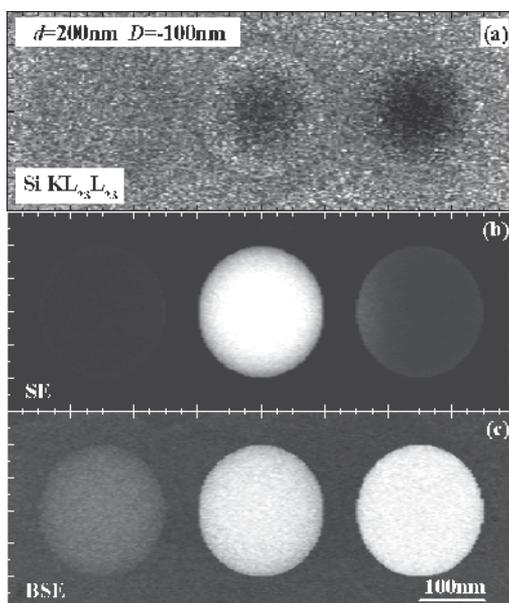


Fig. 39. Calculated SAM- and SEM-images for a specimen made of C-, Cu- and Au-spheres placed on a Si surface and for a primary beam of 6 keV at normal incidence. The diameter of the spheres is 200 nm and the separation between particles is 250 nm: (a) Si(KL<sub>23</sub>L<sub>23</sub>) -Auger electrons; (b) secondary electrons; (c) backscattered electrons.

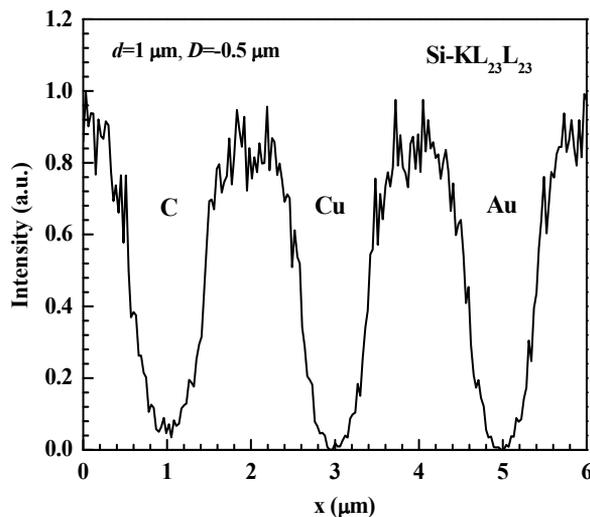


Fig. 40. Calculated line-scan of Si(KL<sub>23</sub>L<sub>23</sub>)-Auger electrons for a specimen made of C-, Cu- and Au-spheres placed on a Si surface and for a primary beam of 6 keV at normal incidence. The diameter of the spheres is 1 μm and the separation between particles is 2 μm .

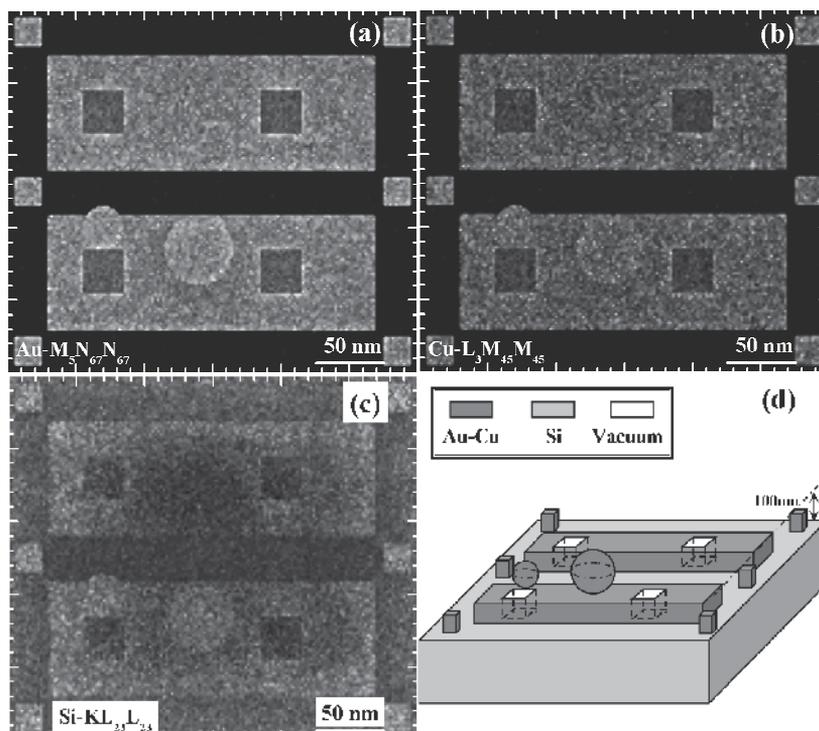


Fig. 41. Simulated Auger electron images for a complex structure of Au-Cu alloy (Au<sub>0.8</sub>Cu<sub>0.2</sub>) on a Si surface and for a primary beam of 6 keV at normal incidence: (a) Au(M<sub>5</sub>N<sub>67</sub>N<sub>67</sub>); (b) Cu(L<sub>3</sub>M<sub>45</sub>M<sub>45</sub>); (c) Si(KL<sub>23</sub>L<sub>23</sub>); (d) sketch map.

In order to demonstrate the universality of the present simulation for complex structures with multi-elemental composition, the simulations of a multi-layer sample filled with Au-Cu alloy has also been performed; the calculated 2D SAM images are shown by Fig. 41. The Au- and Cu-Auger electron images illustrate the specified elemental distribution very well. More interesting is to find that the Si-Auger electron image has a brighter intensity at the position of the geometry features, which should mainly come from the obvious edge effect and backscattering effect for the nanometer system as mentioned above.

In summary, in this section we have proposed a MC simulation method of SAM images for a variety of inhomogeneous specimens made of elements, alloy or compounds with complex geometrical structure. Simulations for several model specimens have been carried out. A good agreement found between the simulation and experimental observation confirms the validity of the present simulation model. The contrast properties for several nanometer structures, typically the particles on a substrate, are discussed in detail by varying particle size and location, energy and incident angle of a primary beam. Several interesting results have been obtained: 1. The depth of particles can influence the contrast dramatically; topography and chemical composition are also the main factors affecting the contrast. 2. The relationships of the contrast of AE images with primary beam energy and particle size have been investigated. The contrast is mainly influenced by different effects, i.e., the edge effect and backscattering effect. 3. At oblique incidence condition the shape of structures observed in a SAM image may be distorted from the realistic one. 4. An enhancement of the substrate Auger signals is mainly due to the stronger backscattering effect of the embedded particle than that of the matrix. 5. The artifact contrast is shown and the backscattering effect of electrons is explained to be the main reason. 6. An effect has been predicted that the substrate Auger signal varies with chemical composition of small particles. 7. Simulation for a multi-layer sample filled with alloy has been performed to show the universality of the simulation method. These results have demonstrated that the present simulation model is universal and useful to explore the contrast mechanism of AE image. Further Monte Carlo studies are necessary in order to develop a correction procedure to remove artifact intensity for a quantitative SAM mapping of realistic surfaces.

## 5. Acknowledgement

This work was supported by the National Natural Science Foundation of China (Grant Nos. 11074232 and 10874160), "973" project (No. 2011CB932801) and "111" project.

## 6. References

- Abe, H.; Hamaguchi, A. & Yamazaki, Y. (2007). Evaluation of CD-SEM measurement uncertainty using secondary electron simulation with charging effect. *Proc. SPIE*, 6518, 65180L-1-10
- Agterveld, D.T.L. van; Palasantzas, G. & Hosson, J.Th.M. De (1999). Surface sensitivity effects with local probe scanning Auger-scanning electron microscopy. *Appl. Phys. Lett.*, 75, 1080-1082
- Asenov, A.; Kaya, S. & Brown, A.R. (2003). Intrinsic parameter fluctuations in decananometer MOSFETs introduced by gate line edge roughness. *IEEE Trans. Elect. Dev.*, 50, 1254-1259

- Ashley, J.C. (1988). Interaction of low-energy electrons with condensed matter: Stopping powers and inelastic mean free paths from optical data. *J. Elect. Spectrosc. Relat. Phenom.*, 46, 199-214
- Ashley, J.C. (1991). Energy loss probabilities for electron, positrons, and protons in condensed matter. *J. Appl. Phys.*, 69, 674-678
- Babin, S; Borisov, S; Miyano, Y; Abe, H; Kadowaki, M; Hamaguchi, A & Yamazaki, Y (2008a). Experiment and simulation of charging effects in SEM. *Proc. SPIE*, 6922, 692219-1-7
- Babin, S; Borisov, S; Ivanchikov, A. & Ruzavin, I. (2008b). Calibration of CD-SEM: moving from relative to absolute measurements. *Proc. SPIE*, 6922, 69222M-1-8
- Berger, M.J. (1963). Monte Carlo calculation of the penetration and diffusion of fast charged particles, in: *Methods in Computational Physics. vol. 1*, Alder, B.; Fernbach, S. & Rotenberg M. (Eds.), 135-215, Academic Press, New York
- Bonham, R.A. & Strand, T.G. (1963). Analytical expression for potential of neutral Thomas-Fermi-Dirac atom and for the corresponding atomic scattering factors for x-rays and electrons. *J. Chem. Phys.*, 39, 2200-2204
- Brandt, W. & Reinheimer, J. (1970). Theory of semiconductor response to charged particles. *Phys. Rev. B*, 2, 3104-3112
- Braun A.E. (2005). *Line Edge Roughness is Here to Stay* (Semiconductor International), Reed Elsevier, New York
- Bronsgeest, M.S.; Barth, J.E.; Swanson, L.W. & Kruit, P. (2008). Probe current, probe size, and the practical brightness for probe forming systems. *J. Vac. Sci. Technol. B*, 26, 949-955
- Bunday, B.D.; Bishop, M.; McCormack, D.; Villarrubia, J.S.; Vladar, A.E.; Dixon, R.; Vorburger, T. & Orji, N.G. (2004). Determination of optimal parameters for CD-SEM measurement of line-edge roughness. *Proc. SPIE*, 5375, 515-533
- Bunday, B. & Allgair, J. (2006). Small feature accuracy challenge for CD-SEM metrology physical model solution. *Proc. SPIE*, 6152, 61520S-1-16
- Bunday, B.; Allgair, J.; Solecky, E.; Archie, C.; Orji, N.G.; Beach, J.; Adan, O.; Peltinov, R.; Bar-zvi, M. & Swyers, J. (2007). The coming of age of tilt CD-SEM. *Proc. SPIE*, 6518, 65181S-1-16
- Bunyan P.J. & Schonfelder J.L. (1965). Polarization by mercury of 100 to 2000 eV electrons. *Proc. Phys. Soc.*, 85, 455-462
- Cailler, M. & Ganachaud, J.P. (1990). Secondary electron emission from solids II. Theoretical description, In: *Fundamental Electron and Ion Beam Interactions with Solids for Microscopy, Microanalysis and Microlithography*, Schou, J.; Kruit, P. & Newbury, D.E. (Eds.), 81-110, Scanning Microscopy Supplement 4, Scanning Microscopy International, Chicago
- Casnati, E.; Tatari, A. & Baraldi, C. (1982). An empirical approach to K-shell ionisation cross section by electrons. *J. Phys. B: At. Mol. Phys.*, 15, 155-168
- Cazaux, J. (1999). Some considerations on the secondary electron emission,  $\delta$ , from e-irradiated insulators. *J. Appl. Phys.*, 85, 1137-1147
- Cazaux, J. (2004). Scenario for time evolution of insulator charging under various focused electron irradiations. *J. Appl. Phys.*, 95, 731-742
- Cazaux, J. (2005). Recent developments and new strategies in scanning electron microscopy. *J. Microsc.*, 217, 16-35
- Cazaux, J. (2006). e-Induced secondary electron emission yield of insulators and charging effects. *Nucl. Instru. Meth. Phys. Res. B*, 244, 307-322

- Choi, Y.; Kim, S. & Han, O. (2006). The CD measuring repeatability enhancement by intensity gradient. *Proc. SPIE*, 6283, 62832F-1-8
- Childs, K.D.; Narum, D.; LaVanier, L.A.; Lindley, P. M.; Schueler, B.W.; Mulholland, G. & Diebold, A.C. (1996). Comparison of submicron particle analysis by Auger electron spectroscopy, time-of-flight secondary ion mass spectrometry, and secondary electron microscopy with energy dispersive x-ray spectroscopy. *J. Vac. Sci. Technol. A*, 14, 2392-2404
- Chung, M.S. & Everhart, T.E. (1977). Role of plasmon decay in secondary electron emission in the nearly-free-electron metals. Application to aluminum. *Phys. Rev. B*, 15, 4699-4715
- Cleary, J.G. & Wyvill, G. (1988). Analysis of an algorithm for fast ray tracing using uniform space subdivision. *The Visual Computer*, 4, 65-83
- Cohen-Tannoudji, C.; Diu, B. & Laloe, F. (1977). *Quantum Mechanics*, Hermann, Paris
- Cowley, J. M. & Liu, J. (1993). Contrast and resolution in REM, SEM and SAM. *Surf. Sci.*, 298, 456-467
- Croon, J.A.; Storms, G.; Winkelmeier, S.; Pollentier, I.; Ercken, M.; Decoutere, S.; Sansen, W. & Maes, H.E. (2002). Line edge roughness: characterization, modeling and impact on device behavior, *Proceedings of IEDM Technical Digest*, pp. 307-310, Leuven Belgium
- Dersch, U.; Korn, A.; Engelmann, C.; Frase, C.G.; Haßler-Grohne, W.; Bosse, H.; Letzkus, F. & Butschke, J. (2005). Impact of EUV mask pattern profile shape on CD measured by CD-SEM. *Proc. SPIE*, 5752, 632-645
- Desalvo, A.; Parisini, A. & Rosa, R. (1984). Monte Carlo simulation of elastic and inelastic scattering of electrons in thin films: I. Valence electron losses. *J. Phys. D.*, 17, 2455-2471
- Desalvo, A. & Rosa, R. (1987). Monte Carlo simulation of elastic and inelastic scattering of electrons in thin films: II. Core electron losses. *J. Phys. D*, 20, 790-795
- Diaz, C.H.; Tao, H.J.; Ku, Y.C.; Yen, A. & Young, K. (2001). An experimentally validated analytical model for gate line-edge roughness (LER) effects on technology scaling. *IEEE Elect. Dev. Lett.*, 22, 287-289
- Ding, Z.J. & Shimizu, R. (1988a). Monte Carlo study of backscattering and secondary electron generation. *Surf. Sci.*, 197, 539-554
- Ding, Z.J.; Shimizu, R.; Sekine, T. & Sakai, Y. (1988b). Theoretical and experimental studies of N(E) spectra in Auger electron spectroscopy. *Appl. Surf. Sci.*, 33/34, 99-106
- Ding, Z.J. & Shimizu, R. (1989a). Theoretical study of the ultimate resolution of SEM. *J. Microsc.*, 154, 193-207
- Ding, Z.J. & Shimizu, R. (1989b). Inelastic collision of kV electrons in solids. *Surf. Sci.*, 222, 313-331
- Ding, Z.J. (1990). *PhD Thesis*, Osaka University
- Ding, Z.J. & Wu, Z.Q. (1993). A comparison of Monte Carlo simulation of electron scattering and x-ray production in solids. *J. Phys. D*, 26, 507-516
- Ding, Z.J.; Shimizu, R. & Goto, K. (1994). Background formation in the low energy region in Auger electron spectroscopy. *J. Appl. Phys.*, 76, 1187-1195
- Ding, Z.J. & Shimizu, R. (1996). A Monte Carlo modeling of electron interaction with solids including cascade secondary electron production. *Scanning*, 18, 92-113
- Ding, Z.J.; Tang, X.D. & Shimizu, R. (2001). Monte Carlo study of secondary electron emission. *J. Appl. Phys.*, 89, 718-726

- Ding, Z.J. & Shimizu, R. (2003). Electron Backscattering and Channelling, In: *Surface Analysis by Auger and X-ray Photoelectron Spectroscopy*, Briggs, D. & Grant J.T. (Eds.), 587-618, IM Publications and SurfaceSpectra Limited, ISBN 1-901019-04-7, Manchester and Chichester
- Ding, Z.J.; Li, H.M.; Tang, X.D. & Shimizu, R. (2004a). Monte Carlo simulation of absolute secondary electron yield of Cu. *Appl. Phys. A*, 78, 585-587
- Ding, Z.J.; Li, H.M.; Goto, K.; Jiang, Y.Z. & Shimizu, R. (2004b). Energy spectra of backscattered electrons in Auger electron spectroscopy: Comparison of Monte Carlo simulation with experiment. *J. Appl. Phys.*, 96, 4598-4606
- Ding, Z.J. & Li, H.M. (2005). Application of Monte Carlo simulation to SEM image contrast of complex structures. *Surf. Interface Anal.*, 37, 912-918.
- Ding, Z.J. & Wang, H.Y. unpublished.
- Drouin, D.; Hovington, P. & Gauvin, R. (1997). CASINO: A new Monte Carlo code in C language for electron beam interactions – part II: Tabulated values of the Mott cross section. *Scanning*, 19, 20-28
- Egerton, R.F. (1986). *Electron Energy Loss Spectroscopy in the Electron Microscope*, Plenum Press, New York
- El Gomati, M.M. & Prutton, M. (1978). Monte Carlo calculations of the spatial resolution in a scanning auger electron microscope. *Surf. Sci.*, 72, 485-494
- El Gomati, M.M.; Janssen, A.P.; Prutton, M. & Venables, J.A. (1979). The interpretation of the spatial resolution of the scanning Auger electron microscope: A theory/experiment comparison. *Surf. Sci.*, 85, 309-316
- El Gomati, M.M.; Prutton, M.; Lamb, B. & Tuppen, C.G. (1988). Edge effects and image contrast in scanning Auger microscopy: A theory/experiment comparison. *Surf. Interface Anal.*, 11, 251-265
- Ercken, M.; Storms, G.; Delvaux, C.; Vandebroek, N.; Leunissien, P. & Pollentier, I. (2002). Line Edge Roughness and Its Increasing Importance, *Proceedings ARCH Interface*
- Fetter, A.L. & Walecka, J.D. (1971). *Quantum Theory of Many-Particle Systems*, McGraw-Hill, New York
- Fink, M. & Yates, A.C. (1970). Theoretical electron scattering amplitudes and spin polarizations. Selected targets, electron energies 100 to 1500 eV. *Atomic Data*, 1, 385-456
- Fink, M. & Ingram, J. (1972). Theoretical electron scattering amplitudes and spin polarizations. Electron energies 100 to 1500 eV. II. Be, N, O, Al, Cl, V, Co, Cu, As, Nb, Ag, Sn, Sb, I, and Ta targets. *Atomic Data*, 4, 129-207
- Forsyth, N.M. & Bean, S. (1994). Low-energy field emission Auger electron spectroscopy. *Surf. Interface Anal.* 22, 338-341
- Foucher, J.; Fabre, A.L. & Gautier, P. (2006). CD-AFM vs CD-SEM for resist LER and LWR measurements. *Proc. SPIE*, 6152, 61520V-1-8
- Frase, C.G. & Haßler-Grohne, W. (2005). Use of Monte Carlo models in the development and validation of CD operators. *Surf. Interface Anal.*, 37, 942-950
- Frase, C.G.; Haßler-Grohne, W.; Dai, G.; Bosse, H.; Novikov, Yu A. & Rakov, A.V. (2007a). SEM linewidth measurements of anisotropically etched silicon structures smaller than 0.1  $\mu\text{m}$ . *Meas. Sci. Technol.*, 18, 439-447
- Frase, C.G.; Buhr, E. & Dirscherl, K. (2007b). CD characterization of nanostructures in SEM metrology. *Meas. Sci. Technol.*, 18, 510-519

- Frase, C.G.; Gnieser, D. & Bosse, H. (2009). Model-based SEM for dimensional metrology tasks in semiconductor and mask industry. *J. Phys. D: Appl. Phys.*, 42, 183001
- Ganachaud, J.P. & Cailler, M. (1979a). A Monte-Carlo calculation of the secondary electron emission of normal metals: I. The model. *Surf. Sci.*, 83, 498-518
- Ganachaud, J.P. & Cailler, M. (1979b). A Monte-Carlo calculation of the secondary electron emission of normal metals: II. Results for aluminium. *Surf. Sci.*, 83, 519-530
- Gauvin, R.; Hovington, P. & Drouin, D. (1995). Quantification of spherical inclusions in the scanning electron microscope using Monte Carlo simulations. *Scanning* 17, 202-219
- Gregory, D. & Fink, M. (1974). Theoretical electron scattering amplitudes and spin polarizations. Electron energies 100 to 1500 eV. III. Li, Na, Mg, P, K, Ca, Sc, Mn, Ga, Br, Sr, Mo, Rh, Cd, Ba, W, and Os targets. *Atomic Data Nucl. Data Tables*, 14, 39-87
- Gorelikov, D.V.; Remillard, J.; Sullioan, N.T. & Davidson, M. (2005). Model-based CD-SEM metrology at low and ultralow landing energies: implementation and results for advanced IC manufacturing. *Surf. Interface Anal.*, 37, 959-965
- Grizinski, M. (1965a). Two-particle collisions. I. General relations for collisions in the laboratory system. *Phys. Rev.*, 138, A305-A321
- Grizinski, M. (1965b). Two-particle collisions. II. Coulomb collisions in the laboratory system of coordinates. *Phys. Rev.*, 138, A322-A335
- Grizinski, M. (1965c). Classical theory of atomic collisions. I. Theory of inelastic collisions. *Phys. Rev.*, 138, A336-A358
- Gwyn, C.W. & Silverman, P.J. (2003). *Photomask Japan*, Yokohama, Japan
- Hagemann, H.J.; Gudat, W. & Kunz, C. (1975). Optical constants from the far infrared to the x-ray region: Mg, Al, Cu, Ag, Au, Bi, C, and Al<sub>2</sub>O<sub>3</sub>. *J. Opt. Soc. Am.*, 65, 742-744
- Hamadeh, E.; Gunther, N.G.; Niemann, D. & Rahman, M. (2006). Gate line edge roughness amplitude and frequency variation effects on intra die MOS device characteristics. *Solid-State Elect.*, 50, 1156-1163
- Hembree, G.G.; Drucker, J.S.; Luo, F.C.H.; Krishnamurthy, M. & Venables, J.A. (1991). Auger electron spectroscopy and microscopy with probe-size limited resolution. *Appl. Phys. Lett.*, 58, 1890-1892
- Hembree, G.G. & Venables, J.A. (1992). Nanometer-resolution scanning Auger electron microscopy. *Ultramicros.*, 47, 109-120
- Hovington, P.; Drouin, D. & Gauvin, R. (1997a). CASINO: A new Monte Carlo code in C language for electron beam interaction -part I: Description of the program. *Scanning*, 19, 1-14
- Hovington, P.; Drouin, D.; Gauvin, R.; Joy, D.C. & Evans, N. (1997b). CASINO: A new Monte Carlo code in C language for electron beam interactions-part III: Stopping power at low energies. *Scanning*, 19, 29-35
- Howell, P.G.T. (1996). A computer program to illustrate macro topography on electron backscattering. *Scanning*, 18, 428-432
- Ichimura, S. (1980). *PhD Thesis*, Osaka University
- Ichimura, S. and Shimizu, R. (1981). Backscattering correction for quantitative Auger analysis. *Surf Sci.*, 112, 386-408
- International Technology Roadmap for Semiconductors (ITRS), Metrology 2007
- Ito, H.; Ito, M.; Magatani, Y. & Soeda, F. (1996). Submicron particle analysis by the Auger microprobe (FE-SAM). *Appl. Surf. Sci.*, 100-101, 152-155
- Jablonski, A. & Powell, C.J. (2005). Monte Carlo simulations of electron transport in solids: applications to electron backscattering from surfaces. *Appl. Surf. Sci.*, 242, 220-235

- Jacka, M. (2001). Scanning Auger microscopy: Recent progress in data analysis and instrumentation. *J. Elect. Spectros. Rela. Phenom.*, 114-116, 277-282
- James, F. (1980). Monte Carlo theory and practice. *Rep. Prog. Phys.*, 43, 1145-1189
- Janssen, A.P. & Venables, J.A. (1978). The effect of backscattered electrons on the resolution of scanning Auger microscopy. *Surf. Sci.*, 77, 351-364
- Jensen, K.O. & Walker, A.B. (1993). Monte Carlo simulation of the transport of fast electrons and positrons in solids. *Surf. Sci.*, 292, 83-97
- Jones, R.; Byers, J. & Conley, W. (2003). Top down versus cross sectional SEM metrology and its impact on lithography simulation calibration. *Proc. SPIE*, 5038, 663-673
- Joy, D.C. (1985). Resolution in low voltage scanning electron microscopy. *J. Microsc.*, 140, 283-292
- Joy, D.C. (1987). A model for calculating secondary and backscattered electron yields. *J. Microsc.*, 147, 51-64
- Joy, D.C. (1995). A database on electron-solid interactions. *Scanning*, 17, 270-275
- Joy, D.C. & Joy, C.S. (1996). Low voltage scanning electron microscopy. *Micron*, 27, 247-263
- Joy, D.C.; Prasad, M.S. & Meyer, H. M. (2004). Experimental secondary electron spectra under SEM conditions. *J. Microsc.*, 215, 77-85
- Kalos, M.H. & Whitlock P.A. (1986). *Monte Carlo Methods. vol. 1*, Wiley, New York
- Kawada, H.; Morokuwa, H.; Takami, S. & Nozoe, M. (2003). CD-SEM for 65-nm process node. *Hitachi Rev.*, 52, 140-146
- Kaya, S.; Brown, A.R.; Asenov, A.; Magot, D. & Linton, T. (2001). Analysis of statistical fluctuations due to line edge roughness in sub-0.1  $\mu\text{m}$  MOSFETs, *Proceedings of SISPAD*, pp. 78-81, Athens, Greece
- Kim, S.D.; Wada, H. & Woo, J.C.S. (2004a). TCAD-based statistical analysis and modeling of gate line-edge roughness effect on nanoscale MOS transistor performance and scaling. *IEEE Trans. Semicond. Manuf.*, 17, 192-200
- Kim, H.W.; Lee, J.Y.; Shin, J.; Woo, S.G.; Cho, H.K. & Moon, J.T. (2004b). Experimental investigation of the impact of LWR on sub-100-nm device performance. *IEEE Trans. Elect. Dev.*, 51, 1984-1987
- Knuth, D.E. (1998). *The Art of Computer Programming, Vol. 3: Sorting and Searching*, Addison-Wesley, Boston, Massachusetts
- Koshikawa, T. & Shimizu, R. (1973). Secondary electron and backscattering measurements for polycrystalline copper with a spherical retarding-field analyser. *J. Phys. D.*, 6, 1369-1380
- Kotera, M. (1989). A Monte Carlo simulation of primary and secondary electron trajectories in a specimen. *J. Appl. Phys.*, 65, 3991-3998
- Kotera, M.; Kishida, T. & Suga, H. (1990). Monte Carlo simulation of secondary electrons in solids and its application for scanning electron microscopy, In: *Fundamental Electron and Ion Beam Interactions with Solids for Microscopy, Microanalysis and Microlithography*, Schou, J.; Kruit, P. & Newbury, D.E. (Eds.), 111-126, Scanning Microscopy Supplement 4, Scanning Microscopy International, Chicago
- Kotera, M.; Yamaguchi, S.; Fujiwara, T. & Suga, H. (1992). Theoretical evaluation of compositional contrast of scanning electron microscope images. *Jpn. J. Appl. Phys.*, 31, 4531-4536
- Li, H.M. & Ding, Z.J. (2005). Monte Carlo simulation of secondary electron and backscattered electron images in scanning electron microscopy for specimen with complex geometric structure. *Scanning*, 27, 254-267

- Li, Y.G.; Mao, S.F., Xiao, S.M. & Ding, Z.J. (2008). Monte Carlo simulation study of SEM images of rough surfaces. *J. Appl. Phys.*, 104, 064901
- Li, Y.G.; Ding, Z.J. & Zhang, Z.M. (2009). Monte Carlo simulation study of scanning Auger electron images. *J. Appl. Phys.*, 106, 024316
- Li, Y.G. (2009). *PhD Thesis*, University of Science and Technology of China
- Lindhard, J. (1954). On the properties of a gas of charged particles. *K. Dan. Vidensk. Selsk. Mat.-Fys. Medd.*, 28, No. 8, 1-57
- Linton, T.; Giles, M. & Packan, P. (1999). The impact of line edge roughness on 100 nm device performance, *Proceedings of Silicon Nanoelectronics Workshop*, pp. 28-29, Tokyo, Japan
- Linton, T.; Chandhok, M.; Rice, B.J. & Schrom, G. (2002). Determination of the line edge roughness specification for 34 nm devices, *Proceedings of IEDM Technical Digest*, pp. 303 - 306, Leuven Belgium
- Liu, J. & Cowley, J.M. (1993). Scanning reflection electron microscopy and associated techniques for surface studies. *Ultramicrosc.*, 48, 381-416
- Liu, J.; Hembree, G.G.; Spinnler, G.E. & Venables, J.A. (1993). Nanometer-resolution surface analysis with Auger electrons. *Ultramicrosc.*, 52, 369-376
- Liu, J. (2000). Contrast of highly dispersed metal nanoparticles in high-resolution secondary electron and backscattered electron images of supported metal catalysts. *Microsc. Microanal.*, 6, 388-399
- Lowney, J.R. (1995). Use of Monte Carlo modeling for interpreting scanning electron microscope linewidth measurements. *Scanning*, 17, 281-286
- Lowney, J.R. (1996). Monte Carlo simulation of scanning electron microscope signals for lithographic metrology. *Scanning*, 18, 301-306
- Luo, S.C.; Zhang, Y.S. & Wu, Z.Q. (1987). A Monte Carlo calculation of secondary electrons emitted from Au, Ag and Cu. *J. Microsc.*, 148, 289-295
- Luo, S.C. & Joy, D.C. (1990). Monte Carlo calculation of secondary electron emission, In: *Fundamental Electron and Ion Beam Interactions with Solids for Microscopy, Microanalysis and Microlithography*, Schou, J.; Kruit, P. & Newbury, D.E. (Eds.), 127-146, Scanning Microscopy Supplement 4, Scanning Microscopy International, Chicago
- Ly, T.D.; Howitt, D.G.; Farrens, M.K. & Harker, A.B. (1995). Monte Carlo calculations for specimens with microstructures. *Scanning*, 17, 220-226
- Maeda, T.; Tanaka, M.; Isawa, M.; Watanabe, K.; Hasegawa, N.; Sekiguchi, K.; Rooyackers, R.; Collaert, N. & Vandeweyer, T. (2008). MuGFET observation and CD measurement by using CD-SEM. *Proc. SPIE*, 6922, 69222P-1-9
- Maksimov, E.G.; Mazin, I.I.; Rashkeev, S.N. & Uspenski, Y.A. (1988). First-principles calculations of the optical properties of metals. *J. Phys. F.*, 18, 833-849
- Mao, S.F.; Li, Y.G.; Zeng, R.G. & Ding, Z.J. (2008). Electron inelastic scattering and secondary electron emission calculated without the single-pole approximation. *J. Appl. Phys.*, 104, 114907
- Matsukawa, T. & Shimizu, R. (1974). A new type of edge effect in high resolution scanning electron microscopy. *Jpn. J. Appl. Phys.*, 13, 583-586
- Matsumoto, J.; Ogiso, Y.; Sekine, M.; Iwai, T. & Whitley, J. (2006). A new algorithm for SEM critical dimension measurements for differentiating between lines and spaces in dense line/space patterns without tone dependence. *Proc. SPIE*, 6349, 634941-1-6

- Mayol, R. & Salvat, F. (1997). Total and transport cross sections for elastic scattering of electrons by atoms. *Atomic Data Nucl. Data Tables*, 65, 55-154
- Metropolis, N. (1987). The beginning of the Monte Carlo method. *Los Alamos Science Special Issue*, 125-130
- Moller, C. (1931). über den Stoss zweier Teilchen unter Berücksichtigung der Retardation der Kräfte. *Z. Phys.*, 70, 786-795
- Morokuma, H.; Miyamoto, A.; Tanaka, M.; Kazui, M. & Takane, A. (2004). New technique to reconstruct effective 3D profile from tilt images of CD-SEM. *Proc. SPIE*, 5375, 727-734
- Mott, N.F. (1929). The scattering of fast electrons by atomic nuclei. *Proc. Roy. Soc. London A*, 124, 425-442
- Mott, N.F. & Massey, H.S.W. (1965). *The Theory of Atomic Collisions*, Oxford University Press, Oxford, UK
- Murata, K.; Kyser, D.F. & Ting, C.H. (1981). Monte Carlo simulation of fast secondary electron production in electron beam resists. *J. Appl. Phys.*, 52, 4396-4405
- Murata, K.; Kawata, H. & Nagami, K. (1987). Electron scattering in low voltage scanning electron microscope targets, In: *Physical Aspects of Microscopic Characterization of Materials*, Kirschner, J.; Murata, K. & Venables, J.A. (Eds.), 83-91, Scanning Microscopy Supplement 1, Scanning Microscopy International, Chicago
- Newbury, D.E.; Myklebust, R.L. & Steel, E.B. (1990). Monte Carlo electron trajectory simulation of x-ray emission from films supported on substrates, In: *Microbeam Analysis-1990*. Michael, J.R. and Ingram, P. (Eds.), 127-130, San Francisco Press, San Francisco
- Nizzoli, F. (1978). A model calculation of the dielectric function in trigonal Se and Te with local-field corrections included. *J. Phys. C.*, 11, 673-683
- Novikov, Yu.A.; Ozerin, Yu.V.; Rakov, A.V. & Todua, P.A. (2007). Method for linear measurements in the nanometre range. *Meas. Sci. Technol.*, 18, 367-374
- Oldiges, P.; Lin, Q.; Petrillo, K.; Sanchez, M.; Jeong, M. & Hargrove, M. (2000). Modeling line edge roughness effect in sub 100 nm gate length devices, *Proceedings of SISPAD*, pp. 131-134, Seattle, WA, USA
- Olson, R.R.; Vanier, L.A. & Narum, D.H. (1993). Backscattering limitations to spatial resolution in the Auger microprobe. *Appl. Surf. Sci.*, 70-71, 266-272
- Palik, E.D. (Ed.) (1985). *Handbook of Optical Constants of Solids*, Academic Press, New York
- Palik, E.D. (Ed.) (1991). *Handbook of Optical Constants of Solids II*, Academic Press, New York
- Penn, D.R. (1987). Electron mean-free-path calculations using a model dielectric function. *Phys. Rev. B*, 35, 482-486
- Pines, D. (1964). *Elementary Excitations in Solids*, Benjamin, New York
- Postek, M.T.; Vldar, A.E.; Lowney, J.R. & Keery, W.J. (2002). Two-dimensional simulation and modeling in scanning electron microscope imaging and metrology research. *Scanning*, 24, 179-185
- Powell, C.J. (1985). Calculation of electron mean free paths from experimental optical data. *Surf. Interface Anal.*, 7, 263-274
- Powell, C.J. (1989). Cross sections for inelastic electron scattering in solids. *Ultramicros.*, 28, 24-31
- Powell, C.J. (2004). Effect of backscattered electrons on the analysis area in scanning Auger microscopy. *Appl. Surf. Sci.*, 230, 327-333
- Prutton, M.; Barkshire, I.R. & Crone, M (1995). Quantitative surface chemical mapping with Auger and backscattered electron signals. *Ultramicros.*, 59, 47-62

- Prutton, M. (2000). From LEED to MULSAM. *Surf. Interface Anal.*, 29, 561-571
- Radzimski, Z.J. & Russ, J.C. (1995). Image simulation using Monte Carlo methods: Electron beam and detector characteristics. *Scanning*, 17, 276-280
- Rao-Sahib T.S. & Wittry D.B. (1974). X-ray continuum from thick elemental targets for 10-50 keV electrons. *J. Appl. Phys.*, 45, 5060-5068
- Rather, H. (1980). *Excitation of Plasmon and Interband Transitions by Electrons*, Springer-Verlag, New York
- Rau, E.I.; Fakhfakh, S.; Andrianov, M.V.; Evstafeva, E.N.; Jbara, O.; Rondot, S. & Mouze, D. (2008). Second crossover energy of insulating materials using stationary electron beam under normal incidence. *Nucl. Instru. Meth. Phys. Res. B*, 266, 719-729
- Reimer, L. & Krefting, E.R. (1976). The effect of scattering models on the results of Monte Carlo simulations, In: *Use of Monte Carlo in Electron Probe Microanalysis and Scanning Electron Microscopy*, Heinrich, K.F.J.; Newbury, D.E. & Yakowitz, H. (Eds.), 45-60, NBS Special Publication 460, US Government Printing Office, Washington, D.C.
- Reimer, L. (1998). *Scanning Electron Microscopy-Physics of Image Formation and Microanalysis, Volume 45 of Spring Series in Optical Sciences, 2nd edition*. Springer, Berlin
- Renoud, R.; Mady, F. Attard, C.; Bigarre, J. & Ganachaud, J.P. (2004). Secondary electron emission of an insulating target induced by a well-focused electron beam -Monte Carlo simulation study. *Phys. Stat. Sol. (a)*, 201, 2119-2133
- Rice, B.J.; Cao, H.; Grumski, M. & Roberts, J. (2006). The limits of CD metrology. *Microele. Eng.*, 83, 1023-1029
- Ritchie, R.H.; Garber, F.W.; Nakai, M.Y. & Birkhoff, R.D. (1969). Low energy electron mean free paths in solids, In: *Advances in Radiation Biology, Vol. 3*, Augenstein, L.G.; Mason, R. & Zelle, M. (Eds.), 1-28, Academic Press, New York
- Rubinstein, R.Y. (1981). *Simulation and the Monte Carlo Method*, Wiley, New York
- Salvat, F.; Fernández-Varea, J. M. & Sempau, J. (2006). PENELOPE-2006: A Code System for Monte Carlo Simulation of Electron and Photon Transport, *Workshop Proceedings*, Barcelona, Spain
- Seah, M.P. & Gilmore, I.S. (1998). Quantitative AES VII. The ionization cross-section in AES. *Surf. Interface Anal.*, 26, 815-824
- Seeger, A.; Fretzagias, C. & Taylo, R. (2003). Software acceleration techniques for the simulation of scanning electron microscope images. *Scanning*, 25, 264-273
- Shiles, E.; Sasaki, T.; Inokuti, M. & Smith, D.Y. (1980). Self-consistency and sum-rule tests in the Kramers-Kronig analysis of optical data: Applications to aluminum. *Phys. Rev. B*, 22, 1612-1628
- Shimizu, R. & Everhart, T.E. (1978). Edge effect in high-resolution scanning Auger-electron microscopy. *Appl. Phys. Lett.*, 33, 549-551
- Shimizu, R. & Ichimura, S. (1981). *Quantitative Analysis by Auger Electron Spectroscopy*, Toyota Foundation Research Rep., Tokyo, Japan
- Shimizu, R. & Ding, Z.J. (1992). Monte Carlo modeling of electron-solid interactions. *Rep. Prog. Phys.*, 55, 487-531
- Shishido, C.; Takagi, Y.; Tanaka, M.; Komuro, O.; Morokuma, H. & Sasada, K. (2002). Characterizing cross-sectional profile variations by using multiple parameters extracted from top-down SEM images. *Proc. SPIE*, 4689, 653-660
- Singhal, S.P. (1975). Dielectric matrix for aluminum. *Phys. Rev. B*, 12, 564-574

- Sramek, S.J. & Cohen, M.L. (1972). Frequency- and wave-vector-dependent dielectric function for Ge, GaAs, and ZnSe. *Phys. Rev. B*, 6, 3800-3804
- Streitwolf, H.W. (1959). Zur Theorie der Sekundärelektronenemission von Metallen der Anregungsprozess. *Ann. Phys. R.*, 3, 183-196
- Sturm, K. (1982). Electron energy loss in simple metals and semiconductors. *Adv. Phys.*, 31, 1-64
- Tanaka, M.; Shishido, C.; Takagi, Y.; Morokuma, H.; Komuro, O. & Mori, H. (2003). Cross-sectional gate feature identification using top-down SEM images. *Proc. SPIE*, 5038, 624-635
- Tanaka, M.; Shishido, C.; Takagi, Y. & Morokuma, H. (2004). MPPC technique for gate etch process monitoring using CD-SEM images and its validity verification. *Proc. SPIE*, 5375, 1144-1155
- Tanaka, M.; Villarubia, J.S. & Vldar, E. (2005). Influence of focus variation on linewidth measurements. *Proc. SPIE*, 5752, 144-155
- Tanaka, M.; Shishido, C. & Kawada, H. (2006). Influence of electron incident angle distribution on CD-SEM linewidth measurements. *Proc. SPIE*, 6152, 61523Z-1-11
- Tanaka, M.; Shishido, C.; Nagatomo, W. & Watanabe, K. (2007). CD-bias evaluation and reduction in CD-SEM linewidth measurements. *Proc. SPIE*, 6518, 651848-1-10
- Tanaka, M.; Shishido, C.; Nagatomo, W. & Watanabe, K. (2008a). Application of model-based library approach to Si<sub>3</sub>N<sub>4</sub> hardmask measurements. *Proc. SPIE*, 6922, 69222L-1-11
- Tanaka, M.; Meessen, J.; Shishido, C. & Watanabe, K. (2008b). CD bias reduction in CD-SEM linewidth measurements for advanced lithography. *Proc. SPIE*, 6922, 69221T-1-11
- Tanuma, S.; Powell, C.J. & Pen, D.R. (1988). Calculation of electron IMFPs for 31 materials. *Surf. Interface Anal.*, 11, 577-589
- Tanuma, S.; Powell, C.J. & Pen, D.R. (2005). Calculations of stopping powers of 100 eV to 30 keV electrons in 10 elemental solids. *Surf. Interface Anal.*, 37, 978-988
- Tokesi, K.; Nemethy, A.; Kover, L.; Varga, D. & Mukoyama, T. (1996). Modeling of electron scattering in thin manganese films on silicon by Monte Carlo methods. *J. Appl. Phys.*, 79, 3763-3769
- Tuppen, C.G. & Davies, G.J. (1985). High spatial resolution Auger linescans across heterogeneous chemical edges by Monte Carlo calculation. *Surf. Interface Anal.*, 7, 235-240
- Umbach, A. & Brünger, W.H. (1989). Spatial resolution tests of scanning Auger microscopy under different topographical conditions. *Surf. Interface Anal.*, 14, 401-413
- Venables, J.A. & Liu, J. (2005). High spatial resolution studies of surfaces and small particles using electron beam techniques. *J. Elect. Spectros. Relat. Phenom.*, 143, 205-218
- Villarrubia, J.S.; Vldar, A.E.; Bunday, B.D. & Bishop, M. (2004). Dimensional metrology of resist lines using a SEM model-based library approach. *Proc. SPIE*, 5375, 199-209
- Villarrubia, J.S.; Vldar, A.E. & Postek, M.T. (2005a). Simulation study of repeatability and bias in the critical dimension scanning electron microscope. *J. Microlithogr. Microfabr. Microsyst.*, 4, 033002
- Villarrubia, J.S.; Vldar, A.E. & Postek, M.T. (2005b). Scanning electron microscope dimensional metrology using a model-based library. *Surf. Interface Anal.*, 37, 951-958
- Villarrubia, J.S. & Ding, Z.J. (2009). Sensitivity of model-based SEM dimensional measurements to model assumptions. *Proc. SPIE*, 7272, 72720R-1-15.; *J. Micro/Nanolith. MEMS MOEMS*, 8, 033003

- Vogel, E.M. (2007). Technology and metrology of new electronic materials and devices. *Nature Nanotech.*, 2, 25-32
- Walker, D.W. (1971). Relativistic effects in low energy electron scattering from atoms. *Adv. Phys.*, 20, 257-323
- Walter, J.P. & Cohen, M.L. (1972). Frequency- and wave-vector-dependent dielectric function for silicon. *Phys Rev B*, 5, 3101-3110
- Wang, H.Y. (2006). *Bachelor Thesis*, University of Science and Technology of China
- Wang, Z.G.; Khuen, S.K.; Fukaya, R.; Kadowaki, Y.; Arai, N.; Ezumi, M. & Satoh, H. (2007). Long-term critical dimension measurement performance for a new mask CD-SEM, S-9380M. *Proc. SPIE*, 6730, 67304T-1-9
- Wells, O. (1974) *Scanning Electron Microscopy*, McGraw-Hill, New York
- Wight, S.A. & Powell, C.J. (2006). Evaluation of the shapes of Auger- and secondary-electron line scans across interfaces with the logistic function. *J. Vac. Sci. Technol. A*, 24, 1024-1210
- Xiong, S. & Bokor, J. (2002). Study of gate line edge roughness effects in 50 nm bulk MOSFET devices. *Proc. SPIE*, 4689, 733-741
- Xiong, S. & Bokor, J. (2004). A simulation study of gate line edge roughness effects on doping profiles of shortchannel MOSFET devices. *IEEE Trans. Elect. Dev.*, 51, 228-232
- Xiong, S.; Bokor, J.; Xiang, Q.; Dudley, I.; Rao, P. & Wang, H.H. (2004). Is gate line edge roughness a first-order issue in affecting the performance of eddp sub-micro bulk MOSFET devices?. *IEEE Trans. Semicond. Manuf.*, 17, 357-361
- Yamaguchi, A.; Tsuchiya, R.; Fukuda, H.; Komuro, O.; Kawada, H. & Iizumi, T. (2003). Characterization of line-edge roughness in resist patterns and estimations of its effect on device performance. *Proc. SPIE*, 5038, 689-698
- Yamaguchi, A.; Ichinose, K.; Shimamoto, S.; Fukuda, H.; Tsuchiya, R.; Ohnishi, K.; Kawada, H. & Iizumi, T. (2004). Metrology of LER: influence of line-edge roughness (LER) on transistor performance. *Proc. SPIE*, 5375, 468-476
- Yamane, T. & Hirano, T. (2005). Sidewall effect of photomask by scanning electron microscope and optical critical dimension metrology. *J. Microlithogr. Microfabr. Microsyst.*, 4, 033003
- Yamazaki, Y. (1977). *PhD Thesis*, Osaka University
- Yan, H.; El Gomati, M.M.; Prutton, M.; Wilkinson, D.K.; Chu, D.P. & Dowsett, M.G. (1998). Mc3D: A three-dimensional Monte Carlo system simulating image contrast in surface analytical scanning electron microscopy I - Object-oriented software design and tests. *Scanning*, 20, 465-484
- Yoshimura, T.; Shiraishi, H.; Yamamoto, J. & Okazaki, S. (1993). Correlation of nano edge roughness in resist patterns with base polymers. *Jpn. J. Appl. Phys*, 32, 6065-6070
- Yue, Y.T.; Li, H.M. & Ding, Z.J. (2005). Monte Carlo simulation of secondary electron and backscattered electron images for a nanoparticle-matrix system. *J. Phys. D: Appl. Phys.*, 38, 1966-1977

# Monte Carlo Simulation of Insulating Gas Avalanche Development

Dengming Xiao

*Dept. of E.E., Shanghai Jiaotong University, 200030  
China*

## 1. Introduction

SF<sub>6</sub> gas has excellent physical and chemical properties: general, chemical stability, not easy to react with other substances; not flammable and explosive, high security and reliability; in particular, to have a very high dielectric strength and destroy arc performance, so the existing power system in gas-insulated equipment, most of them are using pure SF<sub>6</sub> gas as insulating medium. The one of the most serious problem of SF<sub>6</sub> as insulating gas is that the "Kyoto Protocol" on the proposed global warming potential (GWP) in 1997. Its GWP is 23,900 times of CO<sub>2</sub>, at the meeting clearly to reduce the gas as the future or even prohibit the use of the high-pollution gas. Reducing the environmental impact of greenhouse gases has become a research hotspot in recent years.

c-C<sub>4</sub>F<sub>8</sub> is a colorless, odorless, nonflammable gas; and its GWP for the 8700, is one-third of SF<sub>6</sub>, the impact on the environment is far less than the SF<sub>6</sub>; and the gas completely is non-toxic, non-ozone impact. c-C<sub>4</sub>F<sub>8</sub> in the low energy range has a high attachment cross section of in the uniform electric field. The dielectric strength of c-C<sub>4</sub>F<sub>8</sub> is about 1.3 times to the SF<sub>6</sub> gas. c-C<sub>4</sub>F<sub>8</sub> gas used for insulation is existing the shortcomings of expensive, discharge decomposition of conductive particles, and relatively higher liquefaction temperature. By adding N<sub>2</sub>, CO<sub>2</sub> and other buffer gases of inexpensive, lower liquefied temperature and without the existence of carbon decomposition to c-C<sub>4</sub>F<sub>8</sub> gas, c-C<sub>4</sub>F<sub>8</sub> gas discharge characteristics can be improved. Buffer gas through the scattering of the electron energy can reduced electronegative gas to attached the energy range, hindering the avalanche growth, and the dielectric strength of c-C<sub>4</sub>F<sub>8</sub> gas mixtures is reduced less than pure c-C<sub>4</sub>F<sub>8</sub> gas.

c-C<sub>4</sub>F<sub>8</sub> gas mixtures as a dielectric applications has attracted the attention of international experts on electricity and environmental. In 1997 the U.S. Institute of Standards and Technology considered the c-C<sub>4</sub>F<sub>8</sub> gas mixture as a potential future long-term studies insulated gas; 2001 Tokyo Electric Power Industry Center research institutions and the University of Tokyo proposed the application of c-C<sub>4</sub>F<sub>8</sub> gas mixtures as the insulating medium. This article will use the Monte Carlo Simulation (MCS) to study the gas mixtures discharge characteristics of c-C<sub>4</sub>F<sub>8</sub> and N<sub>2</sub>, CO<sub>2</sub> and CF<sub>4</sub> mixed composition.

In the gas discharge, we can not predict the trajectory of one electron, but there are a lot of electron, statistical methods can be used to analyze the randomness of the large number of electron. Avalanche development is effected mostly by the electrical force. From a macro point of view, the electron trajectory is a curve in electric field. Speaking from the micro electronic movement in the gas itself has a random nature. An electron leaved from

emission source, which point in the direction of motion collision was accidental, but has some probability distribution; and gas molecules in multiple collision scattering, the elastic collision in non-electronegative gas, or collisional excitation and impact ionization, may occur. In terms of the electronegative gas, may also occur attached, and there are a variety of collision probability; the energy and direction after the collision has to comply with a certain probability distribution. In electronegative gas, electron may be absorbed by molecular, then the electron movement came to an end, otherwise, continue to the next movement. An electron movement of gas in electric field can be reflected through the collision. Here, necessary to point out that the decision of the location in next collision and the determined direction and energy after the collision are only concerned with this collision, with nothing to do with electron collision before. To suppose that the density of gas molecules is not denseness, there is no chemical reaction between molecules under the conditions, which  $C_4F_8$  gas, and  $N_2$ ,  $CO_2$  and  $CF_4$  gas molecules is satisfied this condition. Random collision occurred in only under the conditions, the Monte Carlo model of establishment and improvement used to calculate the avalanche development of the mixed-gas in this paper.

### **2.1 Monte Carlo simulation model of gas avalanche flow chart**

In this paper, uniform electric field  $E$  in the flat model is introduced. The electron in motion collide with gas molecules, including ionization collisions, excitation collision, elastic collision, which particles produced after the excited state collide also with other electrons and molecules. These processes are random, and the Monte Carlo method can be used to describe these processes. In the space of electronic sample, a certain number (eg 10000) of the simulation electrons sampled, then these electrons state of motion at a given moment statistics is followed and recorded. All electrons are sampled after the electrons moved some time, used to calculate. Figure 2.1 shows the Monte Carlo simulation model of gas discharge avalanche flow chart, followed by the detailed simulation of avalanche development process. Simulation process as follows:

1. sampling a certain number of simulation electrons,
2. calculating the flight time of an electron ,
3. calculating the movement time of the electron in flight time and the change of the location and speed of the electron,
4. deciding the collision type of the electron in flight time,
5. determining the electron state and the direction of movement after occurring the collision of electron,
6. finishing the movement of the electron in given time,
7. repeating the above process for each electron,
8. sampling the velocity, displacement and the number of all electrons in the same sampling interval, to calculate the discharge parameters of the drift velocity, the effective ionization coefficient and so on.

### **2.2 Monte Carlo simulation model of gas avalanche**

#### **2.2.1 Initialization of simulation electrons**

While  $t = 0$ , a large number of electrons of isotropic distribution are released from the emission source, which are the average energy of 1ev (this energy is small enough to not affect the behavior of avalanche). The electrons gain energy under the electric field, and lose energy while it collides with gas molecules. Suppose the density of the gas molecules  $3.32 \times 10^{16} \text{cm}^{-3}$  (i.e. the pressure is 1Torr, temperature is 20 °C), The electron density is small enough to ignore Coulomb effects between electrons.

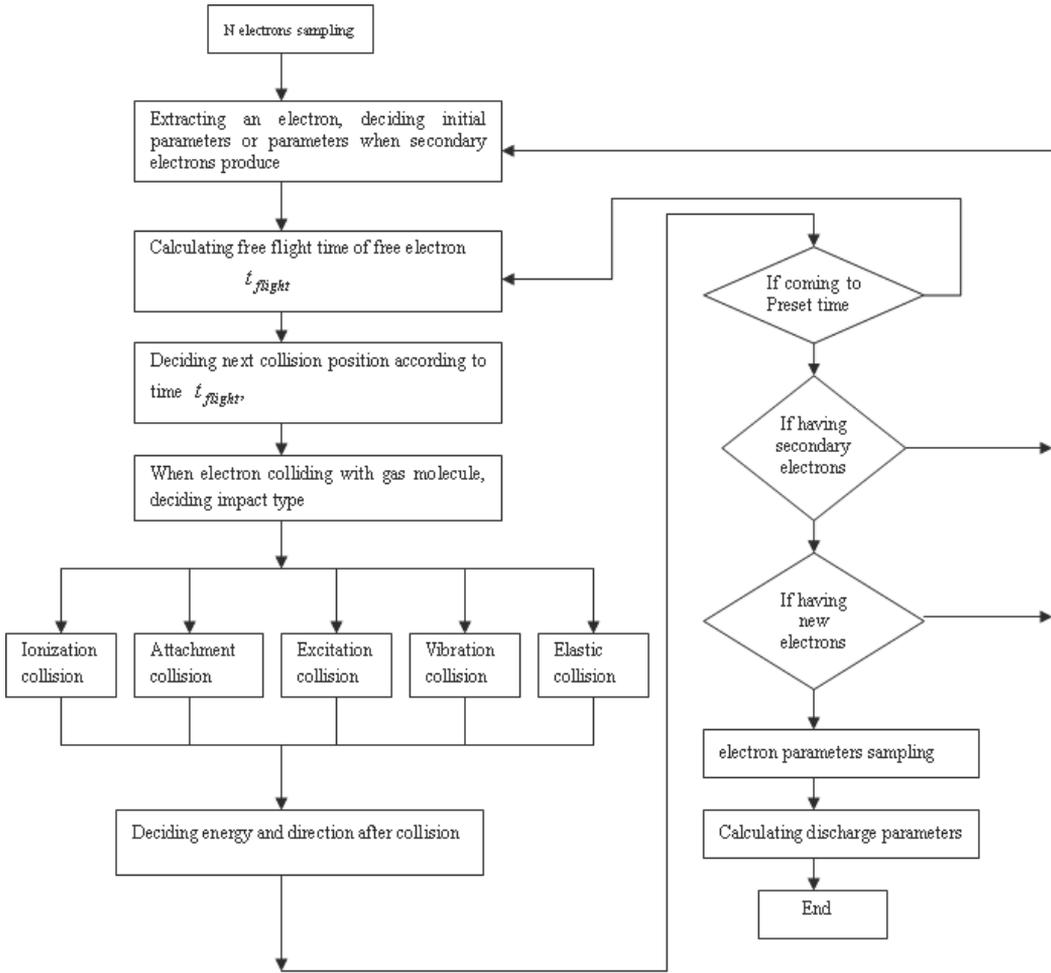


Fig. 2.1 Monte Carlo simulation model of gas discharge avalanche flow chart

**2.2.2 Determination of electronic flight time**

Electronic flight time is calculated in two ways, first using a simple approximation method, using the mean free flight time or the average collision distance, which spent time of a mean free path, another method being randomized to decide.

1. The average collision distance model

In the uniform electric field, the electron moves along the parabola trajectory until a collision with gas molecules. The average collision distance  $\lambda$  is:

$$\lambda = \frac{1}{N\sigma_{total}(\epsilon)} \tag{2.1}$$

Which  $\sigma_{total}(\epsilon)$  is the total collision cross section, in units of  $\text{cm}^2$ , N is the density of gas molecules number, electronic energy  $\epsilon$ .  $\sigma_{total}(\epsilon)$  is a function of energy, obviously  $\lambda$  depends on the electron energy. The mean free path is divided into small segments

$\Delta s = \lambda / k$ . In a step distance, the probability of collision of the electron and gas molecules is:

$$P = 1 - \exp(-N\sigma_{total}(\epsilon)\Delta s) \quad (2.2)$$

Collision occurs or not is determined by a random number  $\xi$ , that is, when P is less than the above calculated, the collision occurred. If no collision, then simulation electrons continue to next step in the movement; if a collision occurs, the number of electron and electron energy changes are determined according to the class of the crash. The whole process will be repeated in the next step until the electron is attached or reach to the set termination conditions.

## 2. The average collision time model

The average collision time of move with drift rate  $W(\epsilon)$  :

$$T_m = \frac{1}{N\sigma_{total}(\epsilon)W(\epsilon)} \quad (2.3)$$

$W(\epsilon)$  is the electron drift velocity.

Used to simulate the average collision time, because a calculation of electron energy changes is so larger that potential error occurs, the actual calculation of the average flight time will be divided into shorter time periods to calculate. Each step is calculated as:

$$dt = \frac{T_m(\epsilon)}{K} \quad (2.4)$$

For large enough K, the collision frequency can be considered in the dt to be constant, then the collision probability in dt is:

$$P = 1 - \exp\left(-\frac{dt}{T_m}\right) \quad (2.5)$$

Using the mean free path or average time of flight, the status of each point in the sequence represents no longer one collision process. A free path or free flight time is divided into a lot of calculation steps, although the accuracy improved, but also causing the increase of computation. For  $c\text{-C}_4\text{F}_8$ ,  $\text{SF}_6$  and other electronegative gas, in order to improve the stability of the simulation, will track more the initial electron. Therefore, previous researchers introduced the concept of free flight time to improve the average flight time method, to enhance the computing speed.

## 3. free flight time

Probability of electron collision in the free flight time  $t_{\text{flight}}$  is:

$$p = f_{total} \cdot t_{\text{flight}} \quad (2.6)$$

In the formula  $f_{total}$  is the total collision frequency, which is a function of time. Set Q (t) is the probability of no collision of electron from 0 to t times, then the probability of no collision from 0 to t +  $t_{\text{flight}}$  time is:

$$Q(t + dt) = Q(t)(1 - P) = Q(t)(1 - f_{total} \cdot t_{\text{flight}}) \quad (2.7)$$

Taking the limit, so  $t_{\text{flight}} \rightarrow 0$ , find:

$$\frac{dQ}{dt} = -f_{\text{total}}(t) \cdot Q(t) \quad (2.8)$$

$$Q(t) = \exp\left[-\int_0^{t_c} f_{\text{total}}(t) dt\right] \quad (2.9)$$

The formula is the probability  $Q(t)$  expression of no collision of the electron from 0 to  $t_c$  moment, thus the probability of collision in  $t_c$  is  $1-Q(t_c)$ . To determine when the collision of the electron occurs is to sample in the division. Producing a [0,1] uniform random number  $\xi$  in the formula and the following represents a uniform distribution random number between 0 and 1:

$$1 - Q(t_c) = \xi \quad (2.10)$$

In principle, the collision time  $t = t_c$  can be obtained from the above two formulas, but then a more complex and time-consuming, because the total collision frequency with time-related will be integrated in the above two formulas. If the total collision frequency is constant, the problem is much better. At this point, it can be set outside the integral sign, then formula (2.10) becomes:

$$t_c = -\frac{\ln(1-\xi)}{f_{\text{total}}} \quad (2.11)$$

As for the  $\xi$  is [0,1] random number, then  $1-\xi$  is [0,1] interval, the formula is changed to:

$$t_c = -\frac{\ln(\xi)}{f_{\text{total}}} \quad (2.12)$$

In fact, the total collision frequency is always time-related, because the formula for calculating the collision frequency :

$$f_{\text{total}} = n_{\text{target}} \sigma(\varepsilon) \cdot v_{\text{relative}} \quad (2.13)$$

It is the function of the electron energy (or speed). Electron move in electric field, and its speed will be changed, and not uniform motion. In order to simplify the calculation of flight time, the null collision method is a good way.

#### 4. null collision

Previous calculation of free flight time, we have seen, if the total electron collision frequency and energy is independent, the calculation of the free flight time will be very simple. In fact electron is accelerated by the field, and its energy is changing, and therefore the total collision frequency will be relation with electron energy, and it will change as the electron energy, the calculation of free flight time can not use the above formula, and will be involved in integration. To deal with this problem is to introduce the concept of null collision, i.e. known as virtual self scattering process by some literatures, and it is a fictitious collision. This process is characterized by its collision frequency so choose - to make the process of the collision and the other collision have the sum of a constant collision frequency. Thus, the total collision frequency is constant, and energy independent. Thus,

when calculating electron collision, it will be very convenient, no matter how the change in electron energy free flight time can be calculated with a formula. Otherwise, having no null-collision case, the electron energy increased significantly in a free flight, then the total collision cross section is to change, the problem of calculation time of the free movement will occur. But without using the above formula, the calculation of free time would complicate. After using the concept of this method is very clear, but when occurring a collision, the collision is not necessarily true, there probably is a real collision, and the collision may be null (no collision). If the collision occurs null, the electron state in motion will remain unchanged, and speed and direction of motion does not change too.

To find the true maximum of the total collision frequency,  $f_{\max} = \max[f_{\text{total}}(\varepsilon)]$ , so the free time of flight can be obtained:

$$t_c = -\frac{\ln(\xi)}{f_{\text{total}}} \quad (2.14)$$

From the above description, the null collision method only requires the total collision frequency to be constant, how to choose it does not affect the calculation of free flight time. However, the selection of null collision frequency directly affect the simulation complexity and speed, and there are many kinds of null impact process on the selection. This article has improved the method, and the traditional and improved methods are described specifically in Section 2.3.

### 2.2.3 determination of the collision point position, the speed and energy

To set the electron affect only by electric fields, the electron in free flight time satisfy the energy conservation law, as the electron affect only by the electric field in the z-axis force, so in  $i+1$  time the speed of the collision points are:

$$v_x^{i+1} = v_x^i \quad (2.15)$$

$$v_y^{i+1} = v_y^i \quad (2.16)$$

$$v_z^{i+1} = v_z^i + \frac{eq}{m} t_{\text{flight}} \quad (2.17)$$

Position and energy of electric field are:

$$x^{i+1} = x^i + v_x^i \cdot t_{\text{flight}} \quad (2.18)$$

$$y^{i+1} = y^i + v_y^i \cdot t_{\text{flight}} \quad (2.19)$$

$$z^{i+1} = z^i + v_z^i \cdot t_{\text{flight}} + \frac{1}{2} \frac{eq}{m} t_{\text{flight}}^2 \quad (2.20)$$

### 2.2.4 Determination of the collision type

In gas discharge, the moving electrons collide with ions and atoms, and the collision type of each process is according to the collision frequency in proportion to the total collision

frequency distribution, which is a random process. The determination of the scattering direction after the collision and speed after scattering is also a random process.

To determine the collision type, the collision of the electron and which kind of gas molecules should be first determined. The collision is the assumption that the density of gas molecules is not big, and there is no chemical reaction between molecules, which occurs only under the conditions of random collisions. According to the statistical physics point of view, the collision probability of the electron and each molecule can first determine, then the collision probability of the electron and various molecules is determine. In this paper, the  $c\text{-C}_4\text{F}_8$  and  $\text{N}_2$ ,  $\text{CO}_2$  and  $\text{CF}_4$  gas molecules in the chemical stability is to satisfy this condition.

Setting the gas make up of the A, B, C three kinds of molecules, the molecular density units (number of molecules in cubic centimeters) are  $N_A$ ,  $N_B$ ,  $N_C$ , then the total frequency of electron collision in gas is:

$$f_t^m(\varepsilon_i) = f_t^A(\varepsilon_i) + f_t^B(\varepsilon_i) + f_t^C(\varepsilon_i) \quad (2.21)$$

Which  $f_t^A, f_t^B, f_t^C, f_t^m$  are respectively the total collision frequency of A, B, C and the mixture gas. They are defined as follows:

$$f_t^A = N_A \sigma_t^A(\varepsilon_i) v \quad (2.22)$$

$$f_t^B = N_B \sigma_t^B(\varepsilon_i) v \quad (2.23)$$

$$f_t^C = N_C \sigma_t^C(\varepsilon_i) v \quad (2.24)$$

As the electron collision frequency is the collision possibility of of electron and molecules, so the probability of collisions of it with A, B, C molecules is respectively:

$$P_A = f_t^A / f_t^m \quad (2.25)$$

$$P_B = f_t^B / f_t^m \quad (2.26)$$

$$P_C = f_t^C / f_t^m \quad (2.27)$$

Easy to determine the collision of electron with which type molecules using standard sampling methods, this article only consider the case of two gas mixtures, then:

$$f_t^m(\varepsilon_i) = f_t^A(\varepsilon_i) + f_t^B(\varepsilon_i) \quad (2.28)$$

Among:

$$f_t^A = N_A \sigma_t^A(\varepsilon_i) v \quad (2.29)$$

$$f_t^B = N_B \sigma_t^B(\varepsilon_i) v \quad (2.30)$$

then the collision probability of two gases are:

$$P_A = f_t^A / f_t^m \quad (2.31)$$

$$P_B = f_t^B / f_t^m \quad (2.32)$$

Determining the collision of electron with which kind of molecular, then the collision type is determined according to the following method.

The collision of electron with gas molecules can occur on impact, such as elastic, scattering, excitation, ionization and attachment, etc.. The total collision frequency of electron and molecular is:

$$f_t = N\sigma_{el}v + N\sigma_{ex}v + N\sigma_i v + N\sigma_a v \quad (2.33)$$

Where  $f_t$  is the total collision frequency,  $\sigma_{el}, \sigma_{ex}, \sigma_i$  and  $\sigma_a$  are respectively the elastic collision cross section, excitation cross sections, ionization cross sections and attachment cross sections.

For the excitation cross section can be further broken down:

$$N\sigma_{ex}v = N\sigma_{ex1}v + N\sigma_{ex2}v + N\sigma_{ex3}v + \dots \quad (2.34)$$

General:

$$N\sigma v = \sum_j N\sigma_j v \quad \sigma = \sum_j \sigma_j \quad (2.35)$$

Where  $j$  is collision types,  $\sigma_{ex1}, \sigma_{ex2}, \sigma_{ex3}, \dots$  are the the collision cross section of different excited levels respectively.

The microscopic cross section of the gas mixtures are represented by  $\sigma_{el}^A, \sigma_{ex}^A, \sigma_i^A$  and  $\sigma_a^A$  (this article used A to represent c-C<sub>4</sub>F<sub>8</sub> in dealing with mixture) and  $\sigma_{el}^B, \sigma_{ex}^B$ , and  $\sigma_i^B$  (B on behalf of non-electronegative gas such as N<sub>2</sub>, CO<sub>2</sub>), said that

$$\sigma_t^A = \sigma_{el}^A + \sigma_{ex}^A + \sigma_i^A + \sigma_a^A \quad (2.36)$$

$$\sigma_t^B = \sigma_{el}^B + \sigma_{ex}^B + \sigma_i^B \quad (2.37)$$

The occurring probability of various collision can be determined as follows.

If the electron and mixed gas molecule A (c-C<sub>4</sub>F<sub>8</sub>) collided, then the occurring probability of various collision are:

$$P_{el}^A = \frac{\sigma_{el}^A}{\sigma_{el}^A + \sigma_{ex}^A + \sigma_i^A + \sigma_a^A} \quad (2.38)$$

$$P_{ex}^A = \frac{\sigma_{ex}^A}{\sigma_{el}^A + \sigma_{ex}^A + \sigma_i^A + \sigma_a^A} \quad (2.39)$$

$$P_i^A = \frac{\sigma_i^A}{\sigma_{el}^A + \sigma_{ex}^A + \sigma_i^A + \sigma_a^A} \quad (2.40)$$

$$P_a^A = \frac{\sigma_a^A}{\sigma_{el}^A + \sigma_{ex}^A + \sigma_i^A + \sigma_a^A} \quad (2.41)$$

If the electron and mixed gas molecule B (other non-electronegative gas molecules) collided, then the occurring probability of various collision are:

$$P_{el}^B = \frac{\sigma_{el}^B}{\sigma_{el}^B + \sigma_{ex}^B + \sigma_i^B} \tag{2.42}$$

$$P_{ex}^B = \frac{\sigma_{ex}^B}{\sigma_{el}^B + \sigma_{ex}^B + \sigma_i^B} \tag{2.43}$$

$$P_i^B = \frac{\sigma_i^B}{\sigma_{el}^B + \sigma_{ex}^B + \sigma_i^B} \tag{2.44}$$

using standard sampling methods, the collision type can be easily determined.

**2.2.5 Determination of electron state after the collision**

After occurring the electron collision, the electron will change the direction of movement (if it will attache to disappear, no longer considered), and the energy is also changing. For different processes, the change is not the same. The energy of elastic collision of electrons does not change almost, but the direction is changed; while ionization occurs, the energy and direction of the electron will be changed, but also creating a new electron; occurring excitation, electron energy and direction of motion will be changed. The random method is used to determine energy, speed and direction after occurring the collision.

1. Determination of the electron scattering angle and azimuth

As the electron energy and direction after collision is related, the electron scattering angle and azimuth are first determined. Scattering angle (or deflexion angle) is the angle of the scattering direction and the direction of the origina motion, and azimuth is angle around the axis of rotation. Typically, the probability density distribution of the azimuth  $\varphi$  is according to  $[0, 2\pi]$  uniform distribution, as long as the  $[0, 1]$  uniform random number  $\xi$  is drawn out, easy to calculate:

$$\varphi = 2\pi\xi \tag{2.45}$$

For the scattering angle  $\chi$ , the produce of random numbers is sampled by the following formula :

$$\frac{2\pi}{\sigma_k(v)} \int_0^\varphi \sigma(v, \chi') \sin \chi' d\chi' = \xi \tag{2.46}$$

Seen from this formula, the scattering angle  $\chi$  is the function of the electron velocity and the random number. Sampling method need to have differential collision cross section, assuming isotropic scattering, which  $\sigma(v, \chi)$  does not rely on  $\chi$ ,  $\sigma_k = 4\pi\sigma$  is gained from this formula, then differential collision cross section is  $\frac{\sigma_k}{4\pi}$ , and the probability of scattering in all directions are the same. At this point, the probability density of the deflection angle  $\chi$

is  $\frac{1}{2}\sin(\chi)$ ,  $\chi \in [0, \pi]$ , and accordingly the isotropic distribution of the scattering angle cosine can be derived:

$$F(x) = 0.5 \times [-1, 1] \quad (2.47)$$

It can be directly sampled, resulting in  $[0, 1]$  random number  $\xi$ , then

$$\cos(\chi) = 2\xi - 1 \quad (2.48)$$

$$\sin(\chi) = \sqrt{1 - \cos^2(\chi)} \quad (2.49)$$

## 2. Determination of the direction cosine after scattering

Determined the direction angle before and after the collision, namely the direction of the scattering angle  $\chi$  and azimuth  $\varphi$ , the direction cosine  $(u_{m+1}, v_{m+1}, w_{m+1})$  after the collision can be determined as follows:

$$u_{m+1} = \frac{(-bcw_m u_m - bdv_m)}{\sqrt{1 - w_m^2}} + au_m \quad (2.50)$$

$$v_{m+1} = \frac{(-bcw_m v_m + bdu_m)}{\sqrt{1 - w_m^2}} + av_m \quad (2.51)$$

$$w_{m+1} = bc\sqrt{1 - w_m^2} + aw_m \quad (2.52)$$

Among

$$a = \cos \chi, \quad b = \sin \chi = \sqrt{1 - a^2} \quad (2.53)$$

$$c = \cos \varphi, \quad d = \sin \varphi \quad (2.54)$$

while  $1 - w_m^2 \rightarrow 0$ , the formula can not be applied, to applied a simple formula:

$$u_{m+1} = bc \quad v_{m+1} = bd \quad w_{m+1} = aw_m \quad (2.55)$$

## 3. Determination of the electron energy and velocity after the collision

If the collision of electron with gas molecules is elastic, the electron energy after collision is:

$$\varepsilon_1 = \varepsilon_0 \left[ 1 - \frac{2m}{M} (1 - \cos \chi) \right] \quad (2.56)$$

Gas molecules are far more weight than electron, so the direction of electron is only changed and the energy of electron does not change basically.

If the excitation process of the electron with gas molecules occurs, then

$$\varepsilon_1 = \varepsilon_0 - \varepsilon_j \quad (2.57)$$

In the formula,  $\varepsilon_1$  is the energy after scattering and the energy  $\varepsilon_0$  for before scattering.  $\varepsilon_j$  is for the energy loss in the excitation process, usually being the excitation energy. The remaining energy will be maintained by the electron.

If the ionization process of the electron with neutral particles occurs, the scattered electron energy and new energy produced secondary electron are distributed by the original remaining electron energy :

$$\varepsilon_{remainder} = \varepsilon_1 - \varepsilon_{ion} \quad (2.58)$$

$$\varepsilon_1' + \varepsilon_2' = \varepsilon_{remainder} \quad (2.59)$$

In the formula, neglected the energy of neutral gas molecules and new ion,  $\varepsilon_1'$  is the electron energy after scattering, and  $\varepsilon_2'$  is the electron energy produced from the ionization, and  $\varepsilon_{ion}$  is the electron energy loss in ionization process. How the two electron energy are allocated? This is correlateion with electron energy differential ionization cross sections  $\sigma(\varepsilon, \varepsilon_c)$ , which can be determined generated a random number  $\xi$  by the following formula, namely to determine the incidence electron collision energy.

$$\xi = \frac{\int_0^{\varepsilon_1'} \sigma(\varepsilon_0, \varepsilon) d\varepsilon}{\sigma_i(\varepsilon_0)} \quad (2.60)$$

In the formula,  $\varepsilon_0$  is the energy before the electron collision, and  $\varepsilon_1'$  is the potential energy of main electron after the collision, and  $\sigma_i(\varepsilon_0)$  is the total ionization cross sections. As the gas ionization energy differential cross sections do not find ready-made data, so the method of random allocation is only introduced. Produced uniform random number  $\xi$  between  $[0,1]$ , then

$$\varepsilon_1' = \varepsilon_{remainder} \xi \quad (2.61)$$

$$\varepsilon_2' = \varepsilon_{remainder} (1 - \xi) \quad (2.62)$$

The size of electron speed can be calculated with energy:

$$v = \sqrt{\frac{2\varepsilon}{m}} \quad (2.63)$$

The velocity component in x, y and z axis can be calculated according to the direction cosine:

$$v_x = v u_{m+1} \quad v_y = v v_{m+1} \quad v_z = v w_{m+1} \quad (2.64)$$

To continue calculating the electron move in the next free flight time, it will be calculated to the set time. After tracking a large number of electron movement, then sampling and calculating electron data, the gas discharge parameters can be got.

### 2.2.6 The sampling and calculation of record data

After the simulation process, the sample processing in the records results is need. The methods of sampling and calculation are different corresponding to different experimental

methods. Here are three kinds of experimental methods involved in sampling and calculation methods.

### 1. SST(Steady State Townsend) experiment

To simulate SST, experimental results must generate a large number of consecutive initial electron, and to form a stable electron flow need to track enough long time, so calculation is exceedingly large amount, then sampling method is:

$$\bar{\omega}(x) = \frac{\sum_{j=1}^N \omega_j \Delta t_j}{\sum_{j=1}^N \Delta t_j} \quad (2.65)$$

Here,  $\omega_j$  is the sampling data of the  $j$  electron in  $x$  to  $x + \Delta x$ , and  $\Delta t_j$  is the time of the electron through the region  $x$  to  $x + \Delta x$ , and  $N$  is the total number of electron in the region, and  $\bar{\omega}(x)$  is the sample data in the location. The method is rarely used because of spending many time.

### 2. TOF(Time Of Flight) experiment

For the TOF experiment, the sampling and calculation are carried out according to the following.

ionization coefficient

$$a = \ln(1 + n_{ion} / N_{i-1}) / (z_i - z_{i-1}) \quad (2.66)$$

attachment

$$\eta = \ln(1 + n_{att} / N_{i-1}) / (z_i - z_{i-1}) \quad (2.67)$$

the drift velocity in avalanche center is defined as

$$W_r = \frac{d\bar{z}}{dt} \quad (2.68)$$

here

$$\bar{z}(t) = \int_{-\infty}^{+\infty} zp(z,t)dz \quad (2.69)$$

$$p(z,t) = \frac{n(z,t)}{\int_{-\infty}^{+\infty} n(z,t)dz} \quad (2.70)$$

here  $n(z, t)$  is the electron density distribution at  $t$  time, then  $W_r$  is defined as:

$$W_r = \frac{(1/N_2) \sum_{j=1}^{N_2} x_j(t_2) - (1/N_1) \sum_{j=1}^{N_1} x_j(t_1)}{t_2 - t_1} \quad (2.71)$$

Axial diffuse coefficient is defined as

$$D_L = \frac{1}{2} \frac{\partial}{\partial t} \int_{-\infty}^{+\infty} [z - \bar{z}(t)]^2 p(z,t) dz \quad (2.72)$$

Therefore

$$D_L = \frac{1}{2} \frac{\left[ \frac{1}{N_2} \sum_{j=1}^{N_2} [z_j(t_2) - \bar{z}_j(t_2)]^2 - \left[ \frac{1}{N_1} \sum_{j=1}^{N_1} [z_j(t_1) - \bar{z}_j(t_1)]^2 \right]}{t_2 - t_1} \right]}{2} \quad (2.73)$$

3. PT(Pulsed Townsend) experiment

For the PT experiment, the electronic properties and its position in the avalanche is correlation, therefore the sampling method conducted by the following

$$\bar{\omega}(t) = \frac{\int_{-\infty}^{+\infty} \int_0^{\infty} \omega F(\varepsilon, x, t) d\varepsilon dx}{\int_{-\infty}^{+\infty} \int_0^{\infty} F(\varepsilon, x, t) d\varepsilon dx} = \frac{\sum_{j=1}^{N_t} \omega_j}{N_t} \quad (2.74)$$

$\omega_j$  is the value of J electron at t time in avalanche, and  $N_t$  is the total number of electron at t time in avalanche, then the avalanche parameters is calculated as follows:

$$W_v = \frac{\bar{z}}{t} \quad (2.75)$$

$$a = \frac{\ln[(n/n_0) + 1]}{\bar{z}} \quad (2.76)$$

The formula and the  $n = n_0 \exp(az - 1)$  in the Townsend equation is the same, in the case of ionizing absence

$$\eta = \frac{n^-}{n^+} \frac{1}{\bar{z}} \quad (2.77)$$

In the case of ionization

$$\eta = \frac{n^-}{n^+} a \quad (2.78)$$

Horizontal and vertical diffusion coefficients have the following definitions

$$D_L(t) = \frac{1}{2t} \frac{d}{dt} \overline{(z_t - \bar{z}_t)^2} = \frac{1}{2t} \frac{d}{dt} \overline{(z_t^2 - \bar{z}_t^2)} \quad (2.79)$$

$$D_r(t) = \frac{1}{2t} \frac{d}{dt} \left\{ \frac{1}{2} \overline{(x^2 + y^2)} \right\} \quad (2.80)$$

Therefore

$$D_L = \frac{\overline{z^2} - \bar{z}^2}{2t} \quad (2.81)$$

$$D_r = \frac{\overline{r^2}}{4t} = \frac{\overline{x^2 + y^2}}{4t} \quad (2.82)$$

Here  $v_e$  is the drift velocity,  $a$  and  $\eta$  are ionization and attachment coefficients respectively, and  $\bar{z}$  is the average distance of motion in a sample model time to the  $t$  time.

### 2.3 Improved Monte Carlo simulation model for the development of avalanche in gas

Contraposing the disadvantage of traditional null collision method, a new Monte Carlo simulation of avalanche development model has been proposed in this section.

While  $t = 0$ , a large number of electrons of the average energy of 1ev isotropic distribution are released from the emission source, the electrons gained energy under the electric field, and they will lose energy as colliding with gas molecules. Suppose the density of the gas molecules is  $3.32 \times 10^{16} \text{cm}^{-3}$  (i.e. the pressure is 1Torr, temperature is 20 °C), The electron density is small enough to ignore the Coulomb interaction between electrons.

1960 S L Lin proposed the null collision Monte Carlo method, to improve its speed, which was cited later by a large number of reference literatures. However the null collision process use the difference of the total collision frequency and the maximum, which citing by the most of author, all cross sections considered must be processed in each step in the simulation process. The null collision method only requires the total collision frequency to be constant, how to choose does not affect the calculation of free flight time. However, the selection of null collision frequency directly affect the simulation complexity and speed, and many kinds of null collision process may be selected. This article is going to set the collision frequency limit for each collision process, then needing only to calculate a cross-section can complete the simulation of gas avalanche development at each interval, to enhance further the calculation speed.

#### 2.3.1 The traditional Monte Carlo null collision method

The traditional way to select a null collision frequency, being cumulative frequency of each collision process, get the real total collision frequency, and find the maximum value  $f_{\max}$  of null collision frequency, and it is the difference of the total collision frequency and the maximum. On this approach, the selected time has two ways: the process set method and pre-set method.

The process set method is to set  $f_{\max}$  in the interval following the experience and according to the electron motion state in each time period, which value is different in different time periods. However, the simulation process should consider two cases, if a free flight time  $f_{\max}$  is always the maximum value of the total collision frequency, then the collision occurred in the end time of free flight time, and the new maximum limit of collision frequency is set in next state, to continue the simulation process. However, as H.R.SKULLERUD pointed out,  $f_{\max}$  is the difficult choice: In principle,  $f_{\max}$  should be made as small as possible to improve computational efficiency in each interval ; but if  $f_{\max}$  made too small, so that  $f_{\text{null}} < 0$  in a certain energy point and the time interval, then the procedure must return to the point of  $f_{\text{null}} = 0$ , to re-select simulation, to makes the code very complex.

Pre-set method is to set the maximum of the total the collision frequency in the range of interest energy before the start of the simulation. As shown in Figure 2.2 ( $\nu_i$  shown the corresponding real frequency in  $i$ -collision process), it pre-calculate the sum of the collision frequency in all energy range of interest, taking the maximum  $f_{\max}$  of sum. This method has the advantage of eliminating the trouble of computing time than the set method, but having

the big null collision frequency, especially in the low energy range, wasting the calculation time of more null collision. The following describe in detail the traditional null collision Monte Carlo simulation process.

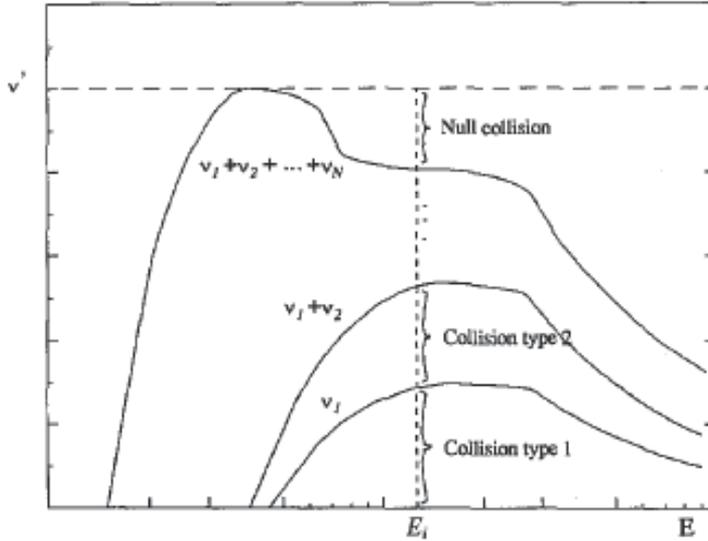


Fig. 2.2 Legend of traditional null collision Monte Carlo method

If describing the electron collision process with  $i=1\dots M$ , in one electron speed, most of the literature used in null collision with  $f_{null}$ :

$$f^i(v) = NQ^i(\epsilon)v \tag{2.83}$$

$$f_{null} = f_{max} - \sum_{i=1}^M f^i(v) \tag{2.84}$$

Where  $N$  is the density of gas molecules,  $Q^i(\epsilon)$  and  $f^i(v)$  are the collision cross section and frequency corresponding to the  $i$  collision process.  $f_{null}$  is the null collision corresponding frequency to no cause any changes in gas properties.  $f_{max}$  is to make always  $f_{null}$  be positive constants, which is the effective collision frequency after introducing null collision. In order to continue the process of simulation, the actual flight time is:

$$t_{flight} = \frac{-1}{f_{max}} \ln(\xi) \tag{2.85}$$

In the formula  $\xi$  and following  $\xi$  express the random number of a uniform distribution between 0 and 1.

Suppose the electron in moving process comply with the law of energy conservation, if to determine the free flight time, while tracking electronic collisions may occur the next time, the displacement and energy can be determined by the classical equation of energy

conservation. Before the next collision, the key is to determine the collision type and the gas molecules possible colliding with the electron.

Collision type will be decided by the random number between a uniform distribution in [0,1]:

$$\xi \leq f_1(v_i) / f_{\max} \quad (2.86)$$

$$f_1(v_i) / f_{\max} < \xi \leq (f_1(v_i) + f_2(v_i)) / f_{\max} \quad (2.87)$$

$$\sum_{j=1}^N f_j(v_i) / f_{\max} < \xi \quad (2.88)$$

If  $\xi$  meets the above equation, the corresponding collision type occurs. For example, when equation (2.88) set up, the null collision type occurs, then the electron does not occur any change.

From the above process can be seen, in order to determine the electron collision frequency and the collision type, all the gas collision cross section must be calculated in each time.

### 2.3.2 Improved null collision Monte Carlo method

In this paper, pre-setting an upper limit  $f_{\max}^i$  for each collision process, the null collision frequency  $f_{null}^i(v)$  corresponding to the  $i$  collision process can be obtained by upper limit  $f_{\max}^i$  and the actual collision frequency  $f^i(v)$ :

$$f_{null}^i(v) = f_{\max}^i - f^i(v) \quad (2.89)$$

The collision frequency unit is  $\text{sec}^{-1}$ , taking into account the density of gas molecules, the ceiling of the total collision frequency will be:

$$f_{\max}^{tot} = \sum_i^M f_{\max}^i \quad (2.90)$$

The actual flight time will be:

$$t_{flight} = \frac{-1}{f_{\max}^{tot}} \ln(\xi) \quad (2.91)$$

First define

$$\xi_i = \frac{1}{f_{\max}^{tot}} \sum_{j=1}^i f_{\max}^j \quad (2.92)$$

if

$$\xi_{i-1} \leq \xi < \xi_i \quad (2.93)$$

Set up, then the  $i$  collision process may occur, the same random number is used to determine whether such a collision is real, if

$$\xi - \xi_{i-1} < \frac{f^{(i)}(v)}{f_{\max}^{tot}} \quad (2.94)$$

Then the collision does occur.

From the above simulation, we can see that the collision type in each collision speed can be determined, only need to calculate a collision cross section.

In the article, the cross-section used are isotropic in the simulation process. in order to reduce the scattering angle, it can be assumed to be isotropic in the laboratory coordinates, to be uniformly distributed in [-1,1] range, so simulation format

$$\cos\theta = 2\xi - 1 \quad (2.95)$$

As the radius of the z-axis is symmetry, azimuth is:

$$\psi = 2\pi\xi \quad (2.96)$$

The energy and direction are determined according to collision type and isotropic angle after the collision.

Simulation in a continuous period of time is until a pre-set time. The energy and location of all electron (including the emergence of new electronic ionization) are sampled at a fixed time interval, then avalanche discharge parameters can be calculated by:

Electron drift velocity  $V_d$  :

$$V_d = \frac{d\bar{z}}{dt} \quad (2.97)$$

Effective ionization coefficient

$$\bar{\alpha} = \frac{\bar{f}}{V_d} = \frac{1}{V_d} \frac{d\ln(n/n_0)}{dt} \quad (2.98)$$

Where  $\bar{z}$  and  $\bar{f}$  are the average distance and the effective ionization frequency in the actual sampling moment of time t.  $n_0$  and  $n$  are the number of electrons while time 0 and t. As  $\bar{\alpha}/N=0$ , the gas dielectric strength  $(E/N)_{\lim}$  can be educed.

## 2.4 The selection of gas collision cross section

Previous Monte Carlo simulation, the cross section area obtained by the experiments are needed all to carry on repeatedly modifying until the cross section area obtained by simulation and the gas discharge test parameters by measurement are consistent. With the development of computer technology, many researchers have adopted neural networks, genetic algorithms and numerical optimization methods to obtain the most data of gas cross section area, to provide a great convenience for future researchers. This article cited the cross section area in the literature to calculate.

In the Monte Carlo calculation, the method to obtain the desirable data are two forms of points cross section and often sub-section cross. In the sub-section, the energy scope of the problem is divided into many intervals, then the section data and energy are independent in each interval, and that is that the section data introduce the sub-section form. Point section refers to tracking particles, starting to find the required data of the each particles by energy

cross section library, then using interpolated method (or otherwise) find the appropriate section of the energy of the various data points, and the approximate formula can be constructed according to variation of characteristics of measured data, then this method is more direct and precise. In this paper, the point section method has been introduced according to known cross-section data.

### 2.4.1 c-C<sub>4</sub>F<sub>8</sub> gas cross section

The measured data of c-C<sub>4</sub>F<sub>8</sub> cross section are many, but more one-sided until 2001 Christophorou and JK Olthoff summarized the cross section of ionization, attachment and neutral decomposition. In 2004, by adjusting the vibration and momentum transformation section according to their basis, Masahiro Yamaji and Yoshiharu Nakamura et al summarized a set c-C<sub>4</sub>F<sub>8</sub> cross-sectional area which Monte Carlo simulation data and measured discharge parameters of drift velocity, vertical diffusion coefficient, effective ionization coefficient are consistent. Their research has given a very accurate c-C<sub>4</sub>F<sub>8</sub> gas cross-sectional area, especially in the low energy range, the cross section shown in Figure 2.3.

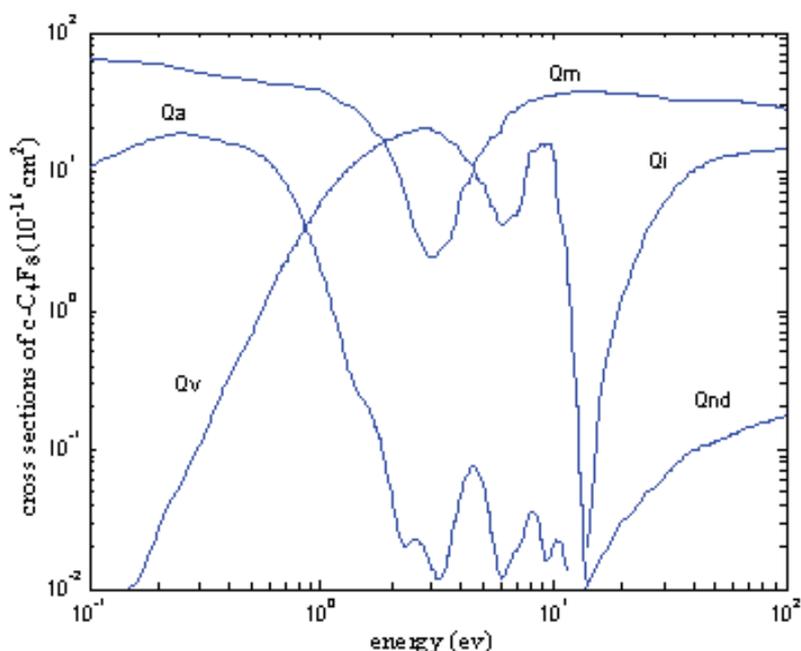


Fig. 2.3 Cross sections of c-C<sub>4</sub>F<sub>8</sub> gas: (Qa), attachment; (Qv), vibration excitation; (Qnd), neutral dissociation; (Qi), ionization; (Qm), momentum transfer.

### 2.4.2 CF<sub>4</sub> gas cross section

This article cited the research data of Kurihara and others in literature, which summarized a set CF<sub>4</sub> cross section area, including 16 kinds of collisions: one elastic momentum transformation, three vibrational excitation and one electronic excitation cross section, one attachment section and seven neutral ionization cross sections and three neutral decomposition section. Kurihara et al have adjusted the decomposed cross-section, making the data and CF<sub>4</sub> gas discharge parameters of the experimental measurement be consistent.

Complete cross-section of the gas as shown in Figure 2.4, due to the corresponding section of the gas are more, Table 2.1 shows the collision process of the gas corresponding to the process ID and the threshold energy.

			Reaction	Threshold (ev)
(1)	$Q_m$	Momentum transformation	$CF_4 + e \rightarrow CF_4 + e$	
(2)	$Q_{v1}$	Vibration excitation	$CF_4 + e \rightarrow CF_4(v1) + e$	0.108
(3)	$Q_{v3}$	Vibration excitation	$CF_4 + e \rightarrow CF_4(v3) + e$	0.168
(4)	$Q_{v4}$	Vibration excitation	$CF_4 + e \rightarrow CF_4(v4) + e$	0.077
(5)	$Q_{ex}$	Electronic excitation	$CF_4 + e \rightarrow CF_4^j + e$	7.54
(6)	$Q_a$	Electronic Attachment	$CF_4 + e \rightarrow F^- + CF_3$	6.4
(7)	$Q_{i1}$	Neutral ionization 1	$CF_4 + e \rightarrow CF_3^+ + F + 2e$	16
(8)	$Q_{i2}$	Neutral ionization 2	$CF_4 + e \rightarrow CF_2^+ + 2F + 2e$	21
(9)	$Q_{i3}$	Neutral ionization 3	$CF_4 + e \rightarrow CF^+ + 3F + 2e$	26
(10)	$Q_{i4}$	Neutral ionization 4	$CF_4 + e \rightarrow C^+ + 4F + 2e$	34
(11)	$Q_{i5}$	Neutral ionization 5	$CF_4 + e \rightarrow F^+ + CF_3 + 2e$	34
(12)	$Q_{i6}$	Neutral ionization 6	$CF_4 + e \rightarrow CF_3^{2+} + F + 3e$	41
(13)	$Q_{i7}$	Neutral ionization 7	$CF_4 + e \rightarrow CF_2^{2+} + 2F + 3e$	42
(14)	$Q_{d1}$	Neutral decomposition 1	$CF_4 + e \rightarrow CF_3 + F + e$	12
(15)	$Q_{d2}$	Neutral decomposition 2	$CF_4 + e \rightarrow CF_2 + 2F + e$	17
(16)	$Q_{d3}$	Neutral decomposition 3	$CF_4 + e \rightarrow CF + 3F + e$	18

Table 2.1.  $CF_4$  gas collisions process and the corresponding threshold energy

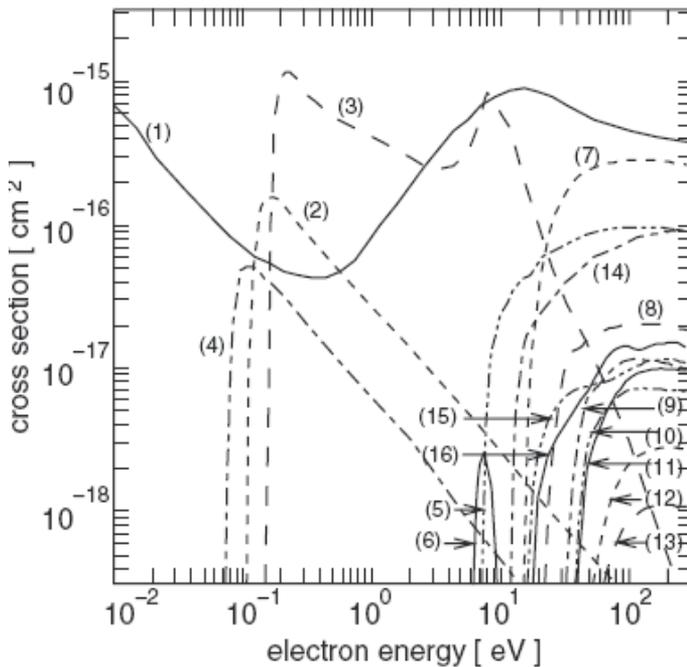


Fig. 2.4  $\text{CF}_4$  gas collision cross section

### 2.4.3 $\text{N}_2$ gas cross section

The  $\text{N}_2$  cross section proposed by AV Phelps and LC Pitchford have been used, and Bolsigplus software package detailed description the section, including one momentum transfer, one rotation, nine vibration excitation, twelve electronic excitation, one single spectrum states, one ionization, considering twenty five of collision cross section. In the software literature describes in more detail.

### 2.4.4 $\text{CO}_2$ gas cross section

Küçükarpacı HN et al scientists have used Monte Carlo simulate the  $\text{CO}_2$  gas avalanche motion in 1979, ordering a set of elastic and inelastic collision section.

### 2.4.5 $\text{SF}_6$ gas cross section

Since  $\text{c-C}_4\text{F}_8$  gas mixtures need comparing with  $\text{SF}_6$  gas mixtures,  $\text{SF}_6$  gas cross section must be chosen. The cross-section sorted out by H. Itoh, Y and others have been used in this article, which including the momentum transformation, vibrational excitation, attachment, ionization and electronic excitation cross sections.

## 2.5 Experimental verification of the improved method

Now  $\text{SF}_6/\text{N}_2$  gas mixtures have been applied in the power industry, however  $\text{SF}_6/\text{CO}_2$  gas mixtures may be superior to certain characteristics of  $\text{SF}_6/\text{N}_2$  mixtures. For example, having the electrode conductive particles between the electrodes and the electrodes rough existing uneven electric field,  $\text{SF}_6-\text{CO}_2$  has a higher insulation breakdown strength.  $\text{SF}_6$  and  $\text{CO}_2$  gas have therefore more reliable collision cross section and the gas discharge experiment data.

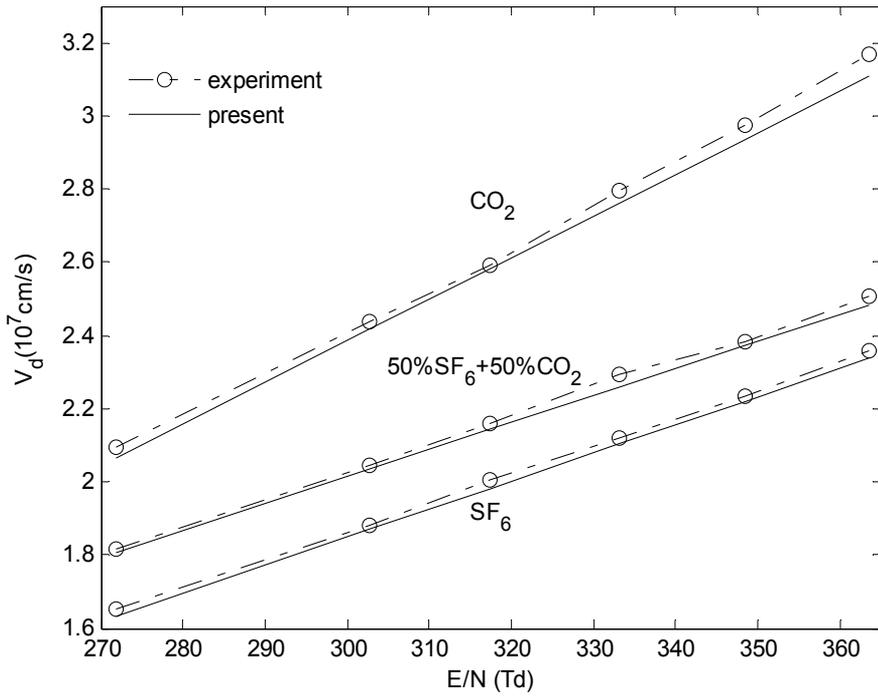


Fig. 2.5 Drift velocity of  $SF_6/CO_2$  gas mixtures

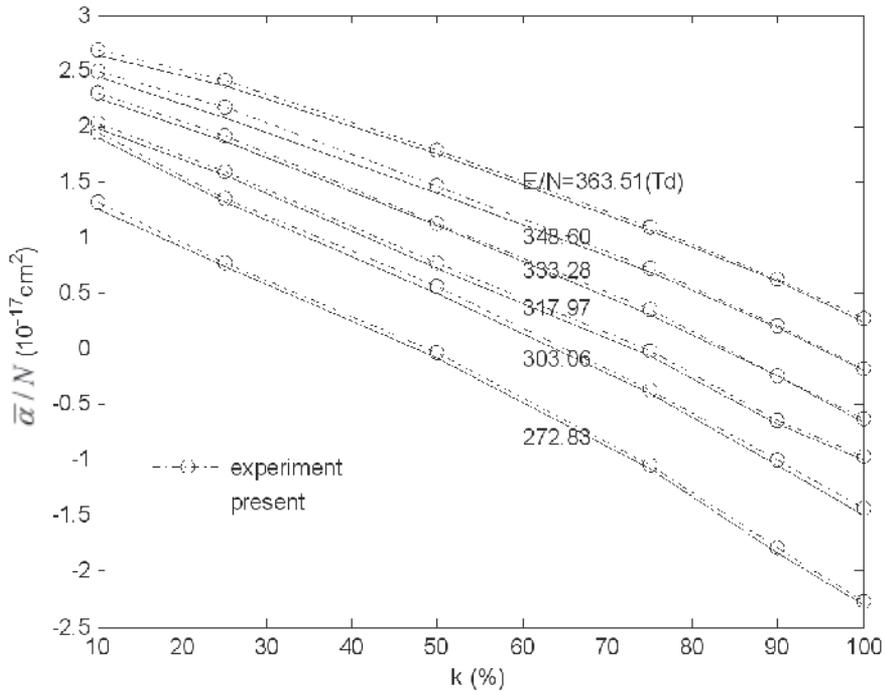


Fig. 2.6  $\bar{\alpha}/N$  of  $SF_6/CO_2$  gas mixtures as a function of mixing ratio  $k$  at different  $E/N$

In order to verify the Monte Carlo method to improve the accuracy, taking the reliable cross-sectional area of  $\text{SF}_6$ ,  $\text{CO}_2$ , the improved Monte Carlo was used simulating the Pulsed Thomson test process of  $\text{SF}_6$ -  $\text{CO}_2$  gas mixtures, to find the drift velocity, effective ionization coefficient of the mixture, to calculate the critical breakdown field strength of the gas mixtures. The simulation results and experimental data were compared to verify the correctness of the modified method.

The electron drift velocity of  $\text{SF}_6$ ,  $\text{CO}_2$  and 50%  $\text{SF}_6$  +50%  $\text{CO}_2$  gas mixtures were shown in figure 2.5. Being seen from the graph,  $\text{CO}_2$  is greater than  $\text{SF}_6$  in the drift speed, and mixed-gas drift velocity is in between, to may be due to  $\text{CO}_2$  gas adding to  $\text{SF}_6$  gas, and the insulation strength of the gas mixtures is reduced. At different field strength  $E / N$ , the variation curves of the effective ionization coefficient  $\bar{\alpha} / N$  as  $\text{SF}_6$  concentration  $k$  was shown in figure 2.6, which decreased with the  $k$  value increases. In figure 2.5 and 2.6 the experimental data were also given to compare with the simulating data, which shows excellent consistency.

The variation of critical breakdown field strength  $(E/N)_{\text{lim}}$  as  $k$  was shown in figure 2.7. The experimental values were also shown in the figure to contrast the simulating data, which showed good agreement, to show that use of the improved Monte Carlo method are feasible.

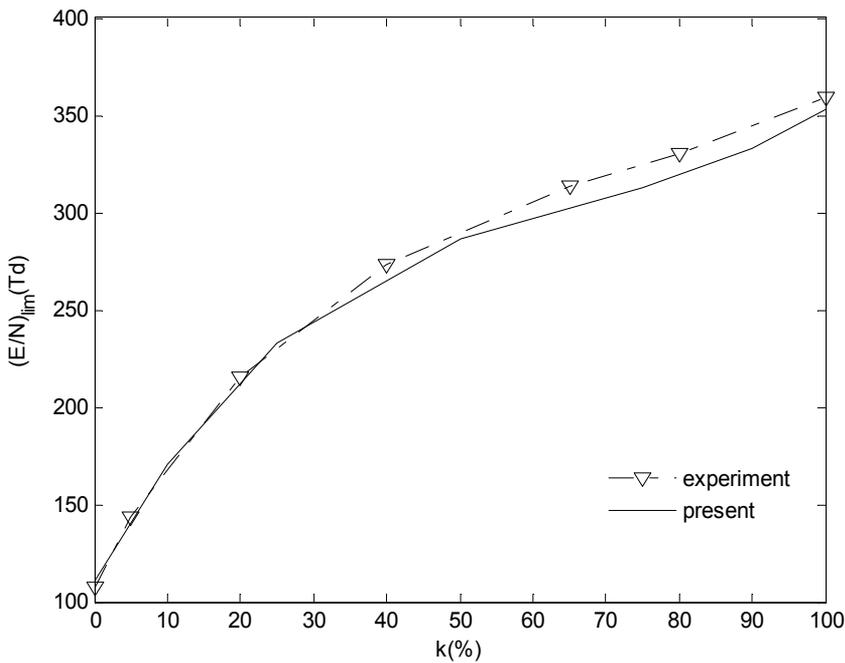


Fig. 2.7 Critical breakdown field strength of  $\text{SF}_6/\text{CO}_2$  mixtures

## 2.6 Simulation of avalanche discharge in c- $\text{C}_4\text{F}_8$ gas mixtures

In this article, the cross section area of gas in reference and the improved Monte Carlo method have been used to simulate the avalanche development process of the c- $\text{C}_4\text{F}_8$  and  $\text{N}_2$ ,  $\text{CO}_2$ ,  $\text{CF}_4$  gas mixtures, to calculate also the critical breakdown field, which were compared with the corresponding data of  $\text{SF}_6$  gas mixtures in the figure. The measuring value of the

critical field strength in the previous literature was also given in the figure to compare with the simulation results to further validate the accuracy of the gas cross section and the improved Monte Carlo method.

Assuming at uniform electric field, the gas dielectric strength increased in proportion with the pressure, then gas pressure increased multiples of the mixture to reach the insulation strength of SF<sub>6</sub> gas was also calculated. At the same time the impact on the environment of the mixed gas used in insulating media, which rate of decline in the GWP, was also analysed.

### 2.6.1 The avalanche discharge parameters of c-C<sub>4</sub>F<sub>8</sub> and N<sub>2</sub> gas mixtures

#### 1. the effective ionization coefficient

The effective ionization coefficient as a function curves of field E/N at different mixture ratio in c-C<sub>4</sub>F<sub>8</sub>/N<sub>2</sub> gas mixtures were shown in Figure 2.8. Shown in the figure,  $\bar{\alpha}/N$  decreased with the increase of c-C<sub>4</sub>F<sub>8</sub> content, increased with the E/N value increases. This is because in the same field strength, c-C<sub>4</sub>F<sub>8</sub> content was higher, then collisions probability of electron and c-C<sub>4</sub>F<sub>8</sub> gas was higher, so the possibility of electron attachment increased to increase attachment coefficient, thus  $\bar{\alpha}/N$  decreased; in the same c-C<sub>4</sub>F<sub>8</sub> content, with the electric field strength increased, high-energy electrons increased, low energy electron decreased, thus  $\bar{\alpha}/N$  increased.

#### 2. electron drift velocity

Simulating drift velocity  $V_e$  of c-C<sub>4</sub>F<sub>8</sub>/N<sub>2</sub> gas mixtures was shown in Figure 2.9. With the c-C<sub>4</sub>F<sub>8</sub> gas content  $k$  increased, at a fixed E/N value, the drift velocity  $V_e$  decreased significantly, which was beneficial from the insulation point of view. The  $V_e$  increased with E/N value increases at a fixed  $K$  value. The  $V_e$  of c-C<sub>4</sub>F<sub>8</sub>/N<sub>2</sub> gas mixtures was affected greatly by c-C<sub>4</sub>F<sub>8</sub> gas.

#### 3. the critical breakdown field strength

To order  $\bar{\alpha}/N = 0$ , the critical breakdown field strength  $(E/N)_{Lim}$  could be given at the different mixing ratio, which showed gas insulation strength due to uniform electric field. Shown from Figure 2.10, SF<sub>6</sub>/N<sub>2</sub> gas mixtures at low mixing ratio, the critical breakdown field strength increased rapidly with SF<sub>6</sub> content  $k$  increase, but larger when mixing ratio, which tended to saturation; and the  $(E/N)_{Lim}$  of c-C<sub>4</sub>F<sub>8</sub>/N<sub>2</sub> gas mixtures increased almost linearly with c-C<sub>4</sub>F<sub>8</sub> content  $k$  increase. In the mixing ratio  $K$  is less than 60%, the insulation strength of c-C<sub>4</sub>F<sub>8</sub>/N<sub>2</sub> was lower than SF<sub>6</sub>/N<sub>2</sub>, but with the mixing ratio increases, the insulation strength of c-C<sub>4</sub>F<sub>8</sub>/N<sub>2</sub> was higher than the insulation strength of SF<sub>6</sub>/N<sub>2</sub>, which was 1.25 times of the latter in the mixing ratio of 100%.

#### 4. the pressure required for the relative insulated intensity of SF<sub>6</sub>

Figure 2.11 showed that c-C<sub>4</sub>F<sub>8</sub>/N<sub>2</sub> and SF<sub>6</sub>/N<sub>2</sub> gas mixtures to reach the insulation strength of SF<sub>6</sub> gas required ratio of pressure and  $K$ . Shown from the figure, the two gas need to increase the pressure of gas mixtures to be similar multiples, when low mixing ratio the former was slightly higher than the latter. With increasing the content of c-C<sub>4</sub>F<sub>8</sub>, gas pressure needed to become smaller and smaller, at  $K=40\%$ , needed to increase 1.35 times, when the mixing ratio continues to increase, the required pressure of c-C<sub>4</sub>F<sub>8</sub>/N<sub>2</sub> and SF<sub>6</sub> was almost similar or even lower than SF<sub>6</sub>.

#### 5. c-C<sub>4</sub>F<sub>8</sub>/N<sub>2</sub> gas mixtures on the improvement of the environmental impact

Figure 2.12 showed the GWP ratio of the gas mixtures and pure SF<sub>6</sub> at the various mixing ratio, in which the greenhouse effect of c-C<sub>4</sub>F<sub>8</sub>/N<sub>2</sub> was far less than SF<sub>6</sub>/N<sub>2</sub>. In 40% of the

mixing ratio, the GWP of SF<sub>6</sub>/N<sub>2</sub> was 40% of the SF<sub>6</sub>, while the GWP of c-C<sub>4</sub>F<sub>8</sub>/N<sub>2</sub> was only 15% of SF<sub>6</sub>, thus environmental impact was reduced.

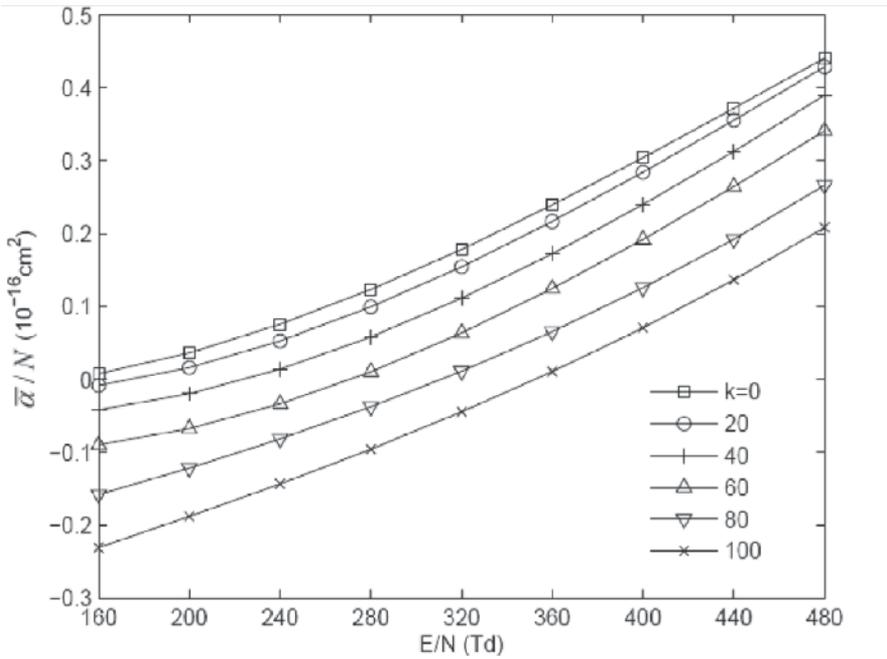


Fig. 2.8  $\bar{\alpha} / N$  of c-C<sub>4</sub>F<sub>8</sub>/N<sub>2</sub> gas mixtures as a function of E / N at different mixture ratio K

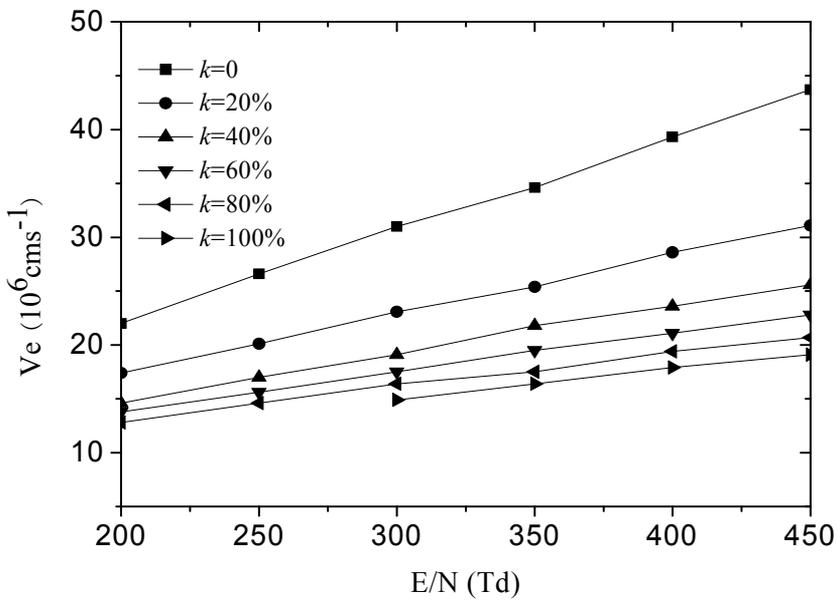


Fig. 2.9 The electron drift velocity  $V_e$  as function of E / N in c-C<sub>4</sub>F<sub>8</sub>/N<sub>2</sub> at different c-C<sub>4</sub>F<sub>8</sub> content  $k$

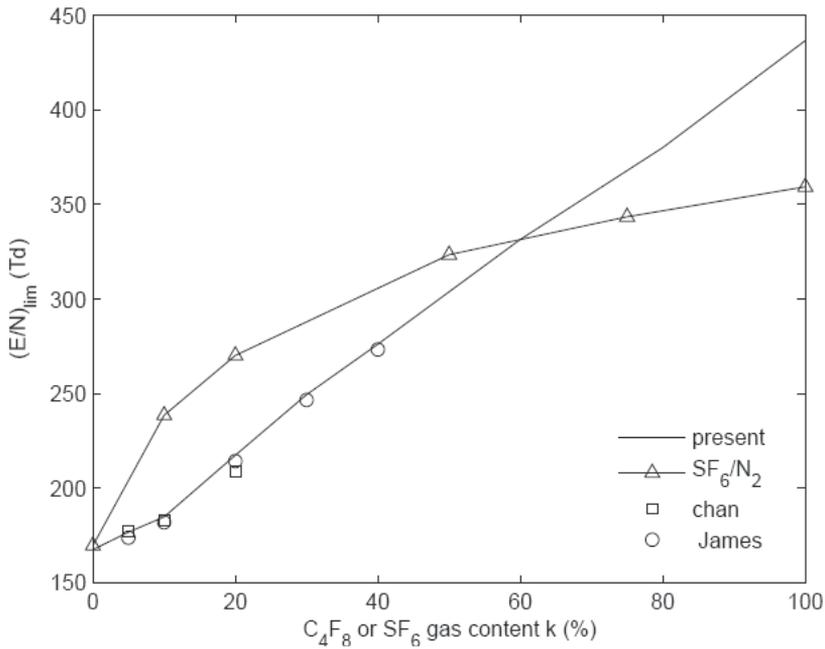


Fig. 2.10  $(E/N)_{lim}$  of  $c-C_4F_8/N_2$  gas mixtures as a function of  $c-C_4F_8$  gas content  $k$  (experiment data:  $SF_6/N_2(\Delta)$ ,  $c-C_4F_8/N_2$  by James( $\circ$ )and Chan ( $\square$ ))

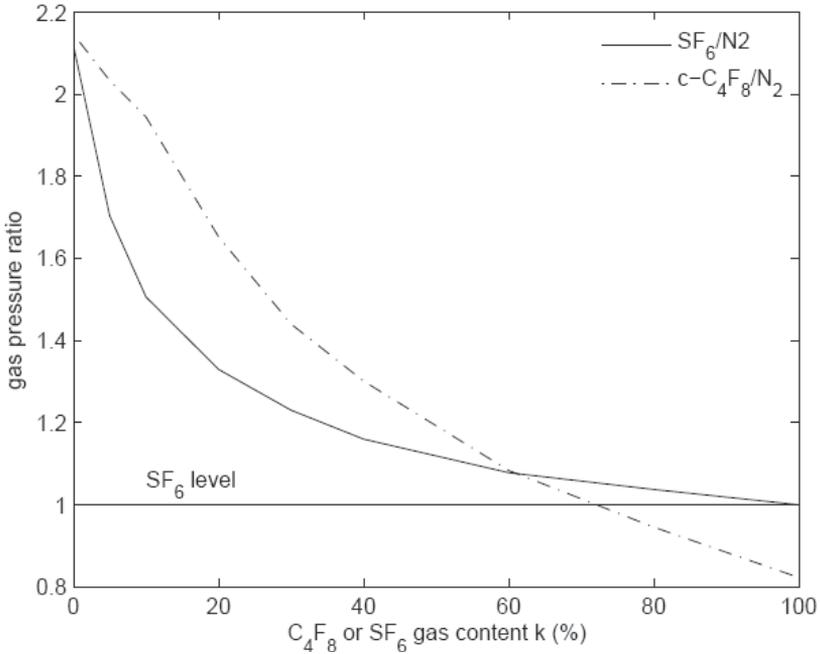


Fig. 2.11 Required gas pressure ratio of  $c-C_4F_8/N_2$  gas mixtures comparable with insulation property of  $SF_6$

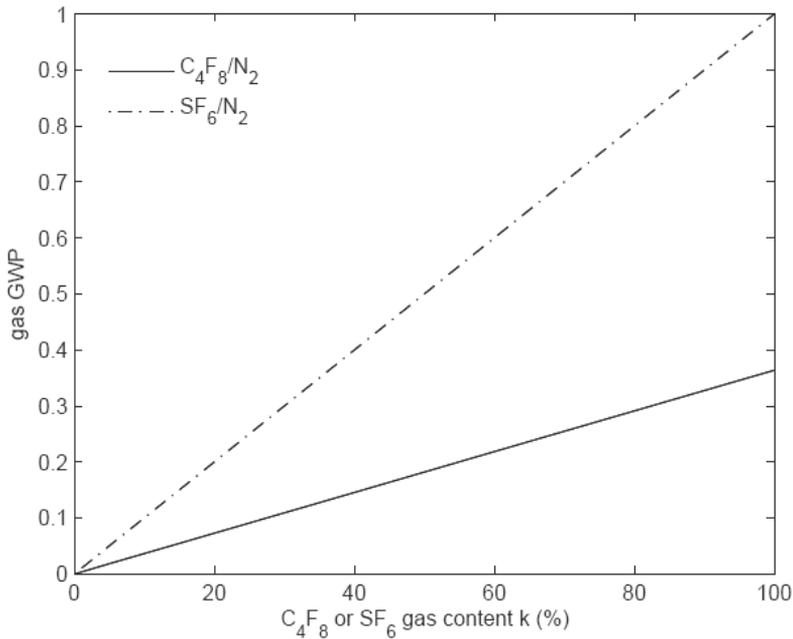


Fig. 2.12 GWP of c-C<sub>4</sub>F<sub>8</sub>/N<sub>2</sub> gas mixtures relative to pure SF<sub>6</sub>

### 2.6.2 The avalanche discharge parameters of c-C<sub>4</sub>F<sub>8</sub> and CO<sub>2</sub> gas mixtures

The effective ionization coefficient  $\bar{\alpha} / N$  in c-C<sub>4</sub>F<sub>8</sub>/CO<sub>2</sub> gas mixtures as a function curve of the field strength E/N at different mixture ratio was shown in Figure 2.13, its variation was almost consistent with c-C<sub>4</sub>F<sub>8</sub>/N<sub>2</sub>.  $\bar{\alpha} / N$  decreased with the increase of c-C<sub>4</sub>F<sub>8</sub> content, and it increased with E/N value increases.

The electron drift velocity  $V_e$  as function of E/N in c-C<sub>4</sub>F<sub>8</sub>/CO<sub>2</sub> mixtures at different c-C<sub>4</sub>F<sub>8</sub> gas mixture ratios  $k$  was shown in Figure 2.14. With the c-C<sub>4</sub>F<sub>8</sub> gas content  $k$  increases, at a fixed E / N value, the drift velocity  $V_e$  trend obvious downward. At  $k$  was same value,  $V_e$  increased with E / N value increases.

The critical breakdown field strength  $(E/N)_{lim}$  in c-C<sub>4</sub>F<sub>8</sub>/CO<sub>2</sub> gas mixtures as a function curve of c-C<sub>4</sub>F<sub>8</sub> gas content  $k$  was shown in Figure 2.15. Shown from the figure, the curve variation was almost consistent comparing SF<sub>6</sub>/CO<sub>2</sub>. While the mixing ratio  $k$  was less than 60%, the  $(E/N)_{lim}$  of c-C<sub>4</sub>F<sub>8</sub>/CO<sub>2</sub> was less than the strength of SF<sub>6</sub>/CO<sub>2</sub>, but with the mixing ratio increases, the insulation strength of c-C<sub>4</sub>F<sub>8</sub>/CO<sub>2</sub> was greater than that of SF<sub>6</sub>/CO<sub>2</sub>.

The ratio of needed pressure of c-C<sub>4</sub>F<sub>8</sub>/CO<sub>2</sub> and SF<sub>6</sub>/CO<sub>2</sub> gas mixtures to reach the insulation strength of SF<sub>6</sub> was shown in Figure 2.16. With the c-C<sub>4</sub>F<sub>8</sub> content increased, the gas pressure needed to become smaller and smaller, at  $k=40\%$ , needed to 1.7 times, and when the mixing ratio continues to increase, the required pressure of c-C<sub>4</sub>F<sub>8</sub>/CO<sub>2</sub> was almost the same with SF<sub>6</sub> gas and SF<sub>6</sub>/CO<sub>2</sub> gas mixtures.

The GWP ratio of c-C<sub>4</sub>F<sub>8</sub>/CO<sub>2</sub>, SF<sub>6</sub>/CO<sub>2</sub> gas mixtures and pure SF<sub>6</sub> at the various mixing ratio was shown in Figure 2.17. The GWP of c-C<sub>4</sub>F<sub>8</sub>/CO<sub>2</sub> was far less than of SF<sub>6</sub>/CO<sub>2</sub>. In 40% of the mixing ratio, the GWP of SF<sub>6</sub>/CO<sub>2</sub> was 40% of GWP for the SF<sub>6</sub>, but the GWP of c-C<sub>4</sub>F<sub>8</sub>/CO<sub>2</sub> was only 15% of GWP in SF<sub>6</sub>, thus impact on the environment was much reduced.

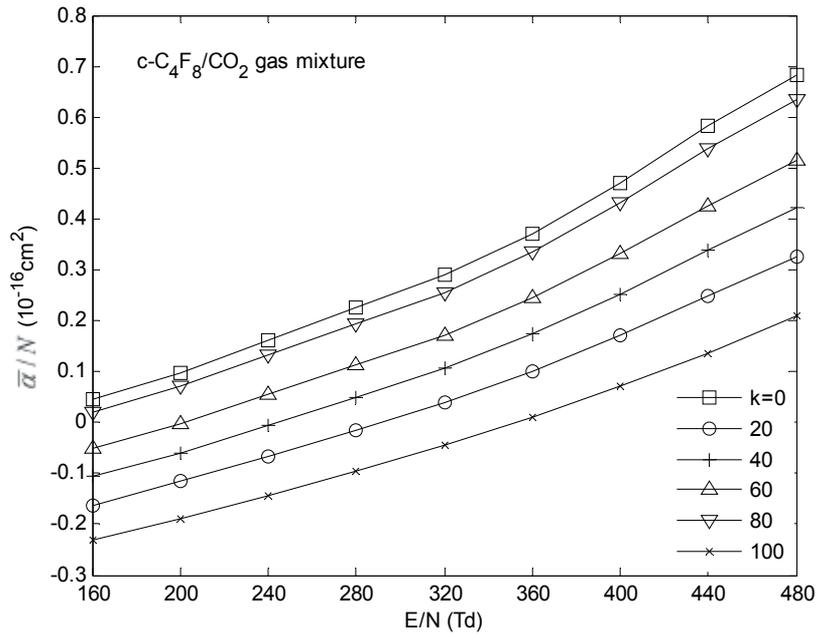


Fig. 2.13  $\bar{\alpha} / N$  in c-C<sub>4</sub>F<sub>8</sub> and CO<sub>2</sub> gas mixtures as a function curve of E/N at different mixture ratio

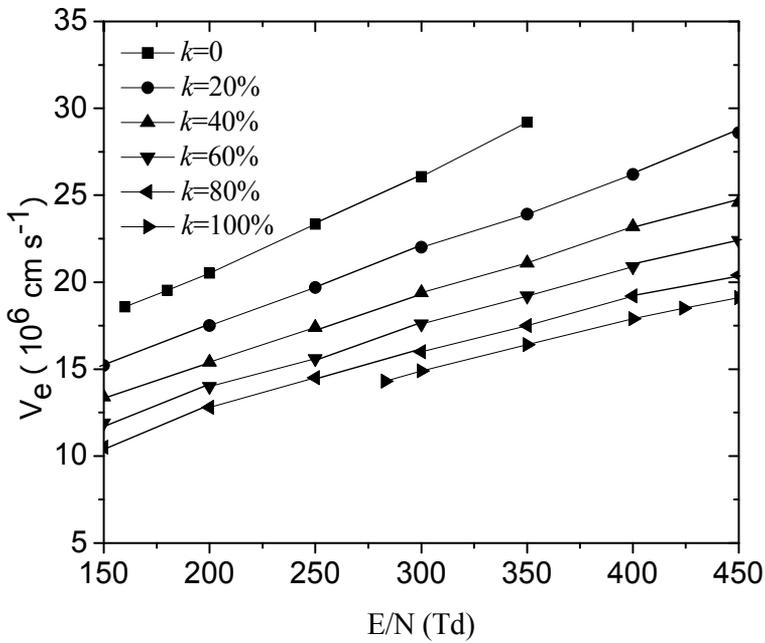


Fig. 2.14 The electron drift velocity  $V_e$  as function of E / N in c-C<sub>4</sub>F<sub>8</sub>/CO<sub>2</sub> mixtures at different c-C<sub>4</sub>F<sub>8</sub> gas mixture ratios k

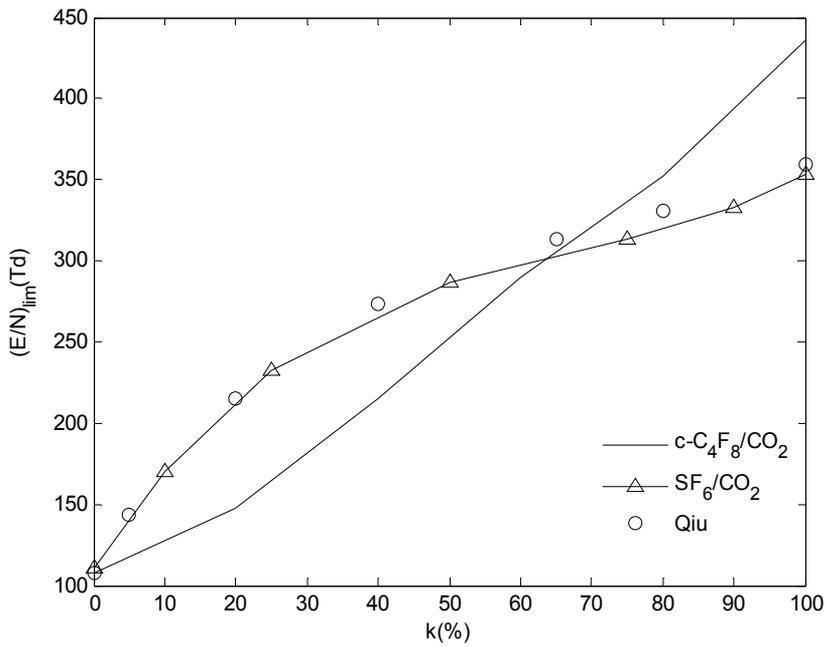


Fig. 2.15  $(E/N)_{lim}$  of  $c-C_4F_8$  and  $CO_2$  gas mixtures as a function of  $c-C_4F_8$  gas content  $k$  (experiment data:  $SF_6/CO_2$  by Qiu)

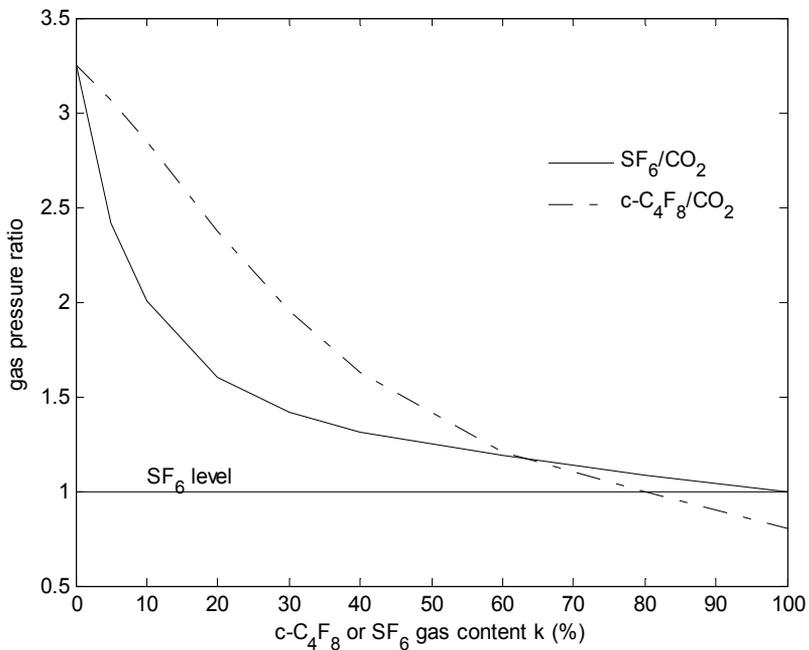


Fig. 2.16 Required gas pressure ratio in  $c-C_4F_8/CO_2$  gas mixtures comparable with insulation property of  $SF_6$

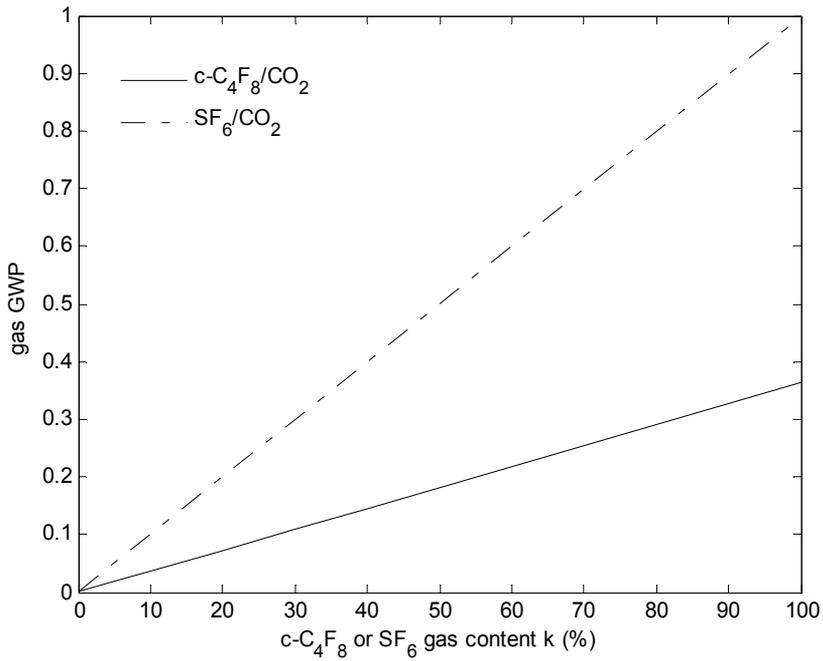


Fig. 2.17 GWP of c-C<sub>4</sub>F<sub>8</sub> and CO<sub>2</sub> gas mixtures relative to pure SF<sub>6</sub>

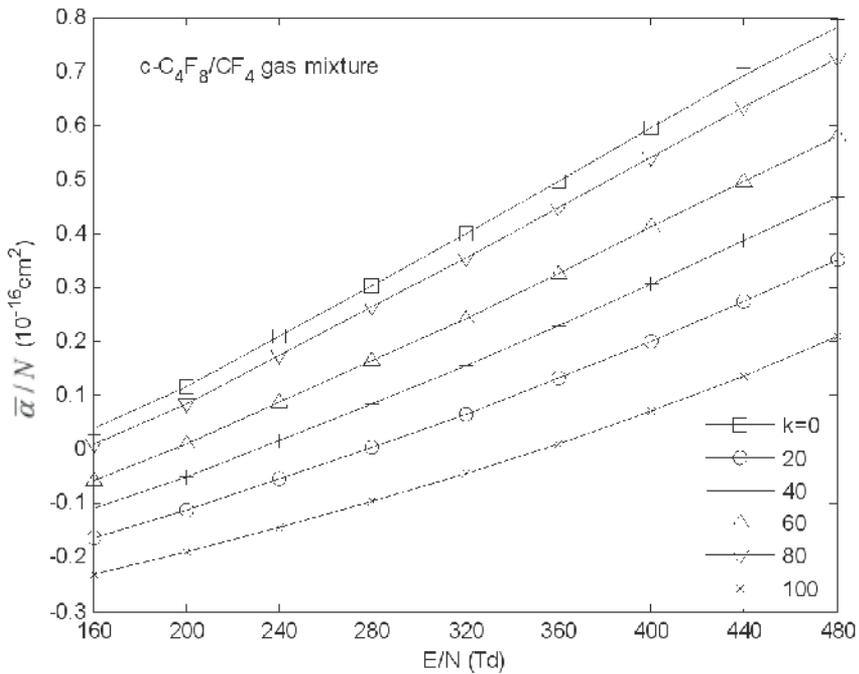


Fig. 2.18  $\bar{\alpha}/N$  of c-C<sub>4</sub>F<sub>8</sub> and CF<sub>4</sub> gas mixtures as a function of E/N at different c-C<sub>4</sub>F<sub>8</sub> gas mixture ratios k

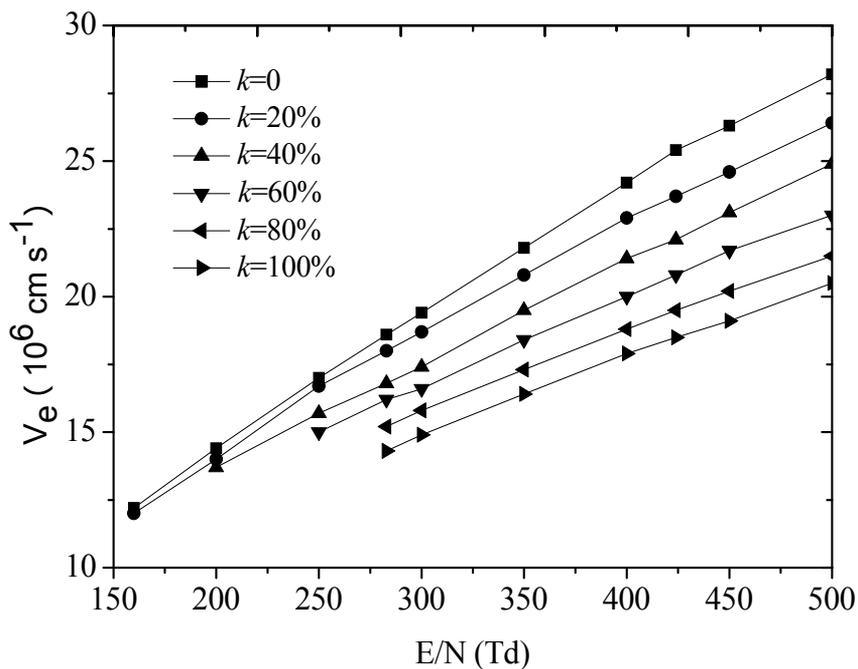


Fig. 2.19 The electron drift velocity  $V_e$  as function of  $E/N$  in  $c\text{-C}_4\text{F}_8/\text{CF}_4$  mixtures at different  $c\text{-C}_4\text{F}_8$  gas mixture ratios  $k$

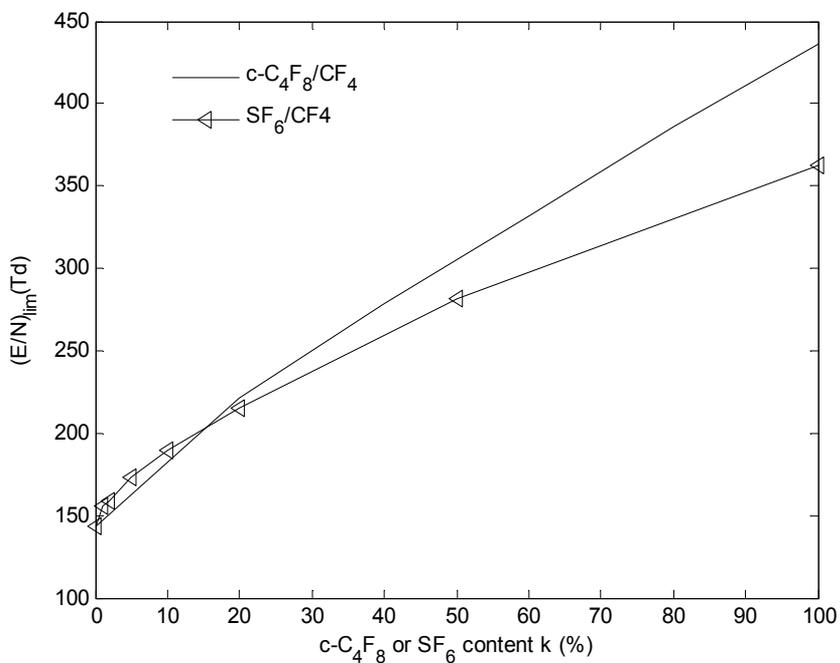


Fig. 2.20  $(E/N)_{\text{lim}}$  of  $c\text{-C}_4\text{F}_8$  and  $\text{CF}_4$  gas mixtures as a function of gas content  $k$

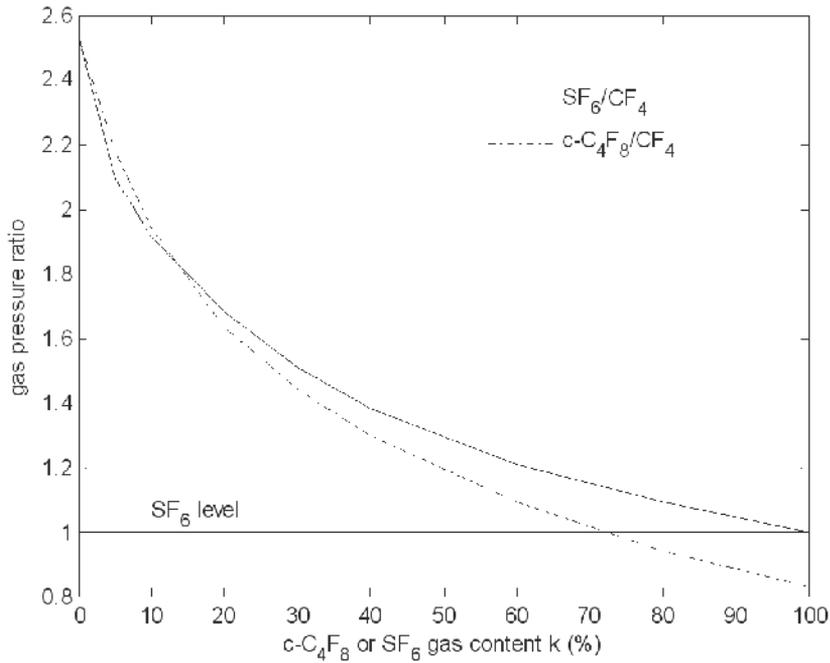


Fig. 2.21 Required gas pressure ratio of c-C<sub>4</sub>F<sub>8</sub> and CF<sub>4</sub> gas mixtures comparable with insulation property of SF<sub>6</sub>

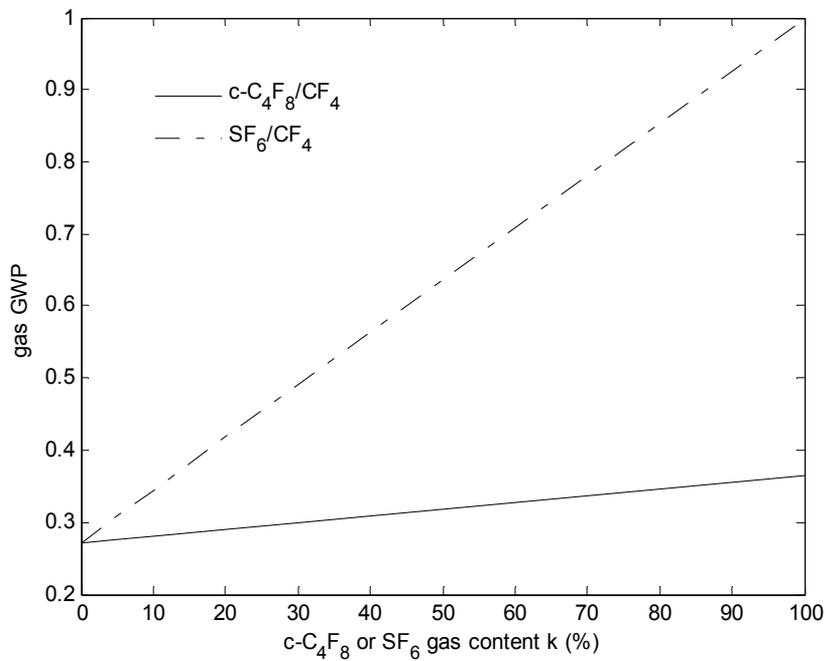


Fig. 2.22 GWP of c-C<sub>4</sub>F<sub>8</sub> and CF<sub>4</sub> gas mixtures relative to pure SF<sub>6</sub>

### 2.6.3 The avalanche discharge parameters of c-C<sub>4</sub>F<sub>8</sub> and CF<sub>4</sub> gas mixtures

The effective ionization coefficient  $\bar{\alpha}/N$  in c-C<sub>4</sub>F<sub>8</sub>/CF<sub>4</sub> changed curve with the field strength at different mixing ratio was shown in Figure 2.18, and  $\bar{\alpha}/N$  decreased with c-C<sub>4</sub>F<sub>8</sub> content increase, but it increased with the E/N value increases.

The electron drift velocity  $V_e$  as function of E/N in c-C<sub>4</sub>F<sub>8</sub>/CF<sub>4</sub> mixtures at different c-C<sub>4</sub>F<sub>8</sub> content  $k$  was shown in Figure 2.19. With the c-C<sub>4</sub>F<sub>8</sub> gas content  $k$  increased, at a fixed E/N value, the drift velocity  $V_e$  trended obvious downward. While  $K$  was at the same,  $V_e$  increased with E/N value increase. The  $V_e$  influenced greatly by CF<sub>4</sub> in c-C<sub>4</sub>F<sub>8</sub>/CF<sub>4</sub> gas mixtures.

The critical breakdown field strength  $(E/N)_{\text{Lim}}$  of c-C<sub>4</sub>F<sub>8</sub>/CF<sub>4</sub> and corresponding SF<sub>6</sub>/CF<sub>4</sub> was compared in Figure 2.20. Although the  $(E/N)_{\text{Lim}}$  of SF<sub>6</sub>/CF<sub>4</sub> increased rapidly at micro-mixing ratio, thus when the corresponding low electronegative gas content, SF<sub>6</sub>/CF<sub>4</sub> dielectric strength slightly higher than c-C<sub>4</sub>F<sub>8</sub>/CF<sub>4</sub>; but with the content  $K$  value increased, the growth rate of the former was not faster than the latter, then with  $K$  increased, the latter was more great than the former.

Figure 2.21 and Figure 2.22 respectively showed the pressure needed the corresponding insulation strength and the GWP in c-C<sub>4</sub>F<sub>8</sub>/CF<sub>4</sub> and SF<sub>6</sub>/CF<sub>4</sub>. To achieve the same insulation strength, the difference of the needed pressure was not great in c-C<sub>4</sub>F<sub>8</sub>/CF<sub>4</sub> and SF<sub>6</sub>/CF<sub>4</sub>. The greenhouse effect of c-C<sub>4</sub>F<sub>8</sub>/CF<sub>4</sub> was far less than that of SF<sub>6</sub>/CF<sub>4</sub>.

## 3. Conclusion

1. The Monte Carlo calculation model to simulate the avalanche development of the gas mixtures has set up, and the flow chart and a detailed simulation steps to simulate gas discharge have also given, to validate the feasibility of their own flight time and using null collision method, to improve the computation speed and stability of average flight distance and average flight time. On the electron moving in gas, the change of the direction of movement, energy and speed after colliding have been analyzed.
2. The improved Monte Carlo simulation method need just to deal with one collision cross section at every step, having a complex cross section for the mixed gas discharge simulation, and then the improved method can greatly improve the calculating speed. Using more mature SF<sub>6</sub>/CO<sub>2</sub> to verify, the improved Monte Carlo method is feasible.
3. The gas collision cross section area used in this article have been detailed described, which cites direct the reference literature.
4. The discharge process in c-C<sub>4</sub>F<sub>8</sub> and N<sub>2</sub>, CO<sub>2</sub>, CF<sub>4</sub> gas mixtures in the uniform electric field have been simulated, to find the effective ionization coefficient, electron drift velocity and the critical breakdown field strength. At the same time the greenhouse effect and the required pressure to achieve insulated strength of SF<sub>6</sub> have been also analyzed, and the corresponding parameters of SF<sub>6</sub> gas mixtures have been compared too.
5. In order to achieve the corresponding insulation strength of SF<sub>6</sub>, the corresponding increase of pressure is similar in c-C<sub>4</sub>F<sub>8</sub> gas mixtures and SF<sub>6</sub> gas mixture. However, the greenhouse effect of c-C<sub>4</sub>F<sub>8</sub> gas mixtures is much less than that of corresponding SF<sub>6</sub> gas mixtures, to reduced much environmental impact.

## 4. Acknowledgment

This work was supported by the National Natural Science Foundation of China (No.: 50777041).

## 5. References

- [1] Takuma T, Gas insulation and greenhouse effect, J IEE Japan 1999,119,232-235
- [2] O.Yamamoto, T.Takuma, S.Hamada and Y.Yamakawa, Applying a Gas Mixture Containing  $c\text{-C}_4\text{F}_8$  as an Insulation medium. IEEE Transactions on Dielectrics and Electrical Insulation, 2001,8(6),1075-1081.
- [3] S L Lin, the null event method in computer simulation, computer physics communications,1978,15,161-163.
- [4] A. SETTAOUTI and L. SETTAOUTI, simulation of electron swarm parameters in  $\text{SF}_6$ , FIZIKA A,2004, 13 (4), 121-136
- [5] V D Stojanović and Z Lj Petrović, Comparison of the results of Monte Carlo simulations with experimental data for electron swarms in  $\text{N}_2$  from moderate to very high electric field to gas density ratios (E/N), J. Phys. D: Appl. Phys,1998,31,834-846.
- [6] Skullerud H R,The stochastic computer simulation of ion motion in a gas subjected to a constant electric field, BRIT.J. APPL. PHYS.1968,2 (1),1567-1568
- [7] R D White, Michael A Morrison and B A Mason, On the use of classical transport analysis to determine cross-sections for low-energy  $e\text{-H}_2$  vibrational excitation, J. Phys. B: At. Mol. Opt. Phys, 2002, 35, 605-626
- [8] W Lowell Morgan, Test of a numerical optimization algorithm for obtaining cross sections for multiple collision processes from electron swarm data, J. Phys. D: Appl. Phys,1993, 26, 209-214.
- [9] Blake Stacey, Relation of Electron Scattering Cross-Sections to Drift. Measurements in Noble Gases [Dissertation], MIT Undergraduate Thesis, 2005
- [10] M. Suzuki et al, Momentum transfer cross section of xenon deduced from electron drift velocity data. J. Phys. D, 1992,25:50-56
- [11] D K Davies, L E Kline and W E Biis, Measurements of swarm parameters and derived electron collision cross sections in methane, J Appl Phys. 1989,65(9):3311-3323.
- [12] L. G. Christophorou and J. K. Olthoff, electron interaction with  $c\text{-C}_4\text{F}_8$ , J. Phys. Chem. Ref. Data, 2001,30, 449-473
- [13] Masahiro Yamaji and Yoshiharu Nakamura, swarm derived electron collision cross section set for the perfluorocyclobutane molecule, J. Phys.D: Appl.Phys,2004,37, 1525-1531
- [14] Kurihara. M, Petrović Z Lj and Makabe T, Transport coefficients and scattering cross-sections for plasma modelling in  $\text{CF}_4\text{-Ar}$  mixtures: a swarm analysis, J. Phys. D:Appl. Phys,2000, 33,2146-2153
- [15] <http://www.codiciel.fr/plateforme/plasma/bolsig/bolsig.php#get>
- [16] A. V. Phelps and L. C. Pitford, Anisotropic scattering of electrons by  $\text{N}_2$  and its effect on electron transport, Phys. Rev. A,1985, 31 2932-2949
- [17] Küçükarpaci H N, Lucas J, Simulation of electron swarm parameters in carbon dioxide and nitrogen for high E/N, J Phy. D: Appl Phys, 1979, 12, 2123-2137
- [18] Qiu Yuchang, Kuffel E, Comparison of  $\text{SF}_6/\text{N}_2$  and  $\text{SF}_6/\text{CO}_2$  Gas Mixtures as Alternative to  $\text{SF}_6$  Gas, IEEE Trans on DEI, 1999, 6 (6) : 892- 895.

- [19] Xiao D M, Li X G, Xu X. Swarm parameters in SF<sub>6</sub> and CO<sub>2</sub> gas mixtures, J. Phys. D: Appl. Phys. 2001, 34: L133-L135
- [20] Chan C C, Pace M and Christophorou L G, Gaseous Dielectrics I,1980, 11 ,149
- [21] J de Urquijo, E Basurto and J L Herná'ndez-A'vila, Measurement of electron drift, diffusion, and effective ionization coefficients in the SF<sub>6</sub>-CHF<sub>3</sub> and SF<sub>6</sub>-CF<sub>4</sub> gas mixtures, J. Phys. D: Appl. Phys.,2003, 36, 3132-3137

# Monte Carlo Simulation of Electron Dynamics in Doped Semiconductors Driven by Electric Fields: Harmonic Generation, Hot-Carrier Noise and Spin Relaxation

Dominique Persano Adorno  
*Dipartimento di Fisica e Tecnologie Relative,  
Università di Palermo and CNISM-INFN,  
Viale delle Scienze, edificio 18, I-90128 Palermo  
Italy*

## 1. Introduction

In solid state electronics the miniaturization of integrated circuits implies that, even at moderate applied voltages, the components can be exposed to very intense electric fields. Advances in electronics push the devices to operate also under cyclostationary conditions, i.e. under large-signal and time-periodic conditions. A main consequence of this fact is that circuits exhibit a strongly nonlinear behavior. Furthermore, semiconductor based devices are always imbedded into a noisy environment that could strongly affect their performance, setting the lower limit for signal detection in electronic circuits. For this reason, to fully understand the complex scenario of the nonlinear phenomena involved in the devices response, an analysis of the electron dynamics in low-doped semiconductors far from equilibrium conditions is very important. Semiconductor spintronics offers a possible direction towards the development of hybrid devices that could perform logic operations, communication and storage, within the same material technology: electron spin could be used to store information, which could be transferred as attached to mobile carriers and, finally, detected. Despite these advantages, for the operability of prospective spintronic devices, the features of spin relaxation at relatively high temperatures, jointly with the influence of transport conditions, should be firstly well understood.

This chapter reviews recent results obtained by using a three dimensional semiclassical multivalleys Monte Carlo code to simulate the nonlinear carrier dynamics in low-doped GaAs bulks (Persano Adorno et al., 2009a; Persano Adorno, 2010; Spezia et al., 2010). The aim is to discuss and clarify the most relevant findings obtained by the investigation of: (a) the harmonic generation process and (b) the spectral density of the electron velocity fluctuations in the presence of intense sub-terahertz electric fields;(c) the influence of temperature and transport conditions on the electron spin relaxation.

Since Monte Carlo approach includes, at a microscopic level, all the sources of the nonlinearities (hot carriers, velocity overshoot, intervalley transfer, etc.) which take place in electronic devices operating under large-signal conditions, it allows to study harmonic

generation in the presence of far-infrared fields. Research along this line is motivated both by the desire to understand the response of doped semiconductors to high-power submillimeter radiation and by the possibility of applications to frequency conversion of coherent radiation in the THz frequency region, where other conventional techniques are ineffective or difficult to realize. In the first part of the chapter, the results of the analysis of the polarization of the generated harmonics are reported (Persano Adorno, 2010). In particular, the polarization obtained from the mixing of an oscillating electric field and a static field is compared with that obtained in the presence of two cyclostationary fields, having an integer ratio between the two frequencies. The findings show that the strength and the polarization of the mixed-fields emission exhibit a strong dependence on the angle between the orientation of the two fields. Unusual polarization features of the generated harmonics are found and discussed.

Recently, studies concerning the constructive aspects of noise and fluctuations in different non-linear systems have shown that the addition of external noise to systems with an intrinsic noise may result in a less noisy response. In the central part of the chapter the attention is focused on the calculation of the modifications of the electronic noise spectra caused by the addition of an external correlated noise source. Numerical results show that, under specific conditions, the presence of a fluctuating component, added to a driving oscillating electric field, can reduce the total noise power. Furthermore, a non-linear behavior of the spectral density with the noise intensity is found. Critically depending on the external noise correlation time, the system benefits from the constructive interplay between the random fluctuating electric field and the intrinsic noise of the system. This is a relevant example of noise enhanced stability (NES) in semiconductor systems (Persano Adorno et al., 2009a).

The last part of the chapter is dedicated to the investigation of the influence of temperature and transport conditions on the electron spin depolarization, making use of a Monte Carlo code which includes the precession description of the spin polarization vector. In order to make spintronics a viable prospective technology it needs to find out the best conditions to achieve long spin relaxation times and/or spin diffusion lengths in semiconductor materials. Electron spin depolarization lengths and times show a nonmonotonic dependence on both the lattice temperature and the electric field amplitude (Spezia et al., 2010). Both parameters appear to be fundamental for the design and fabrication of spintronic devices.

## 2. Monte Carlo approach

The correct theoretical description of a semiconductor device can be obtained by self-consistently solving the Boltzmann transport equation, or the quantum mechanical equivalent of it, and the Maxwell field equations. In the presence of high amplitude driving fields or under cyclostationary conditions, no analytical solution of the Boltzmann equation is known. Approximate solutions can be obtained within drift-diffusion or hydrodynamic models, but the validity of these models is very limited. It becomes necessary to perform a numerical simulation of the process and Monte Carlo approach represents one of the most powerful methods to simulate the transport properties in semiconductor devices, beyond the quasi-equilibrium approximations. This technique, representing a space-time continuous solution of the field and transport equations, is suitable for studying both the steady state and the dynamic characteristics of the devices. Owing to its flexibility, Monte Carlo method presents the remarkable advantage of giving a detailed description of the particle motion in the semiconductor, by taking into account the main details of band structure, scattering processes and heating effects, specific device design and material parameters. It allows to obtain important electron dynamics information, such as the average velocity,

temperature, current density, etc., directly without the need of calculating first the electron distribution function. The time interval between two collisions (time of free flight), the scattering mechanisms, the collisional angle, and all parameters of the problem are chosen in a stochastic way, making a mapping between the probability density of the given microscopic process and a uniform distribution of random numbers.

In our code the conduction bands of GaAs are represented by the  $\Gamma$ -valley, by four equivalent L valleys and three X-valleys. The algorithm includes: (i) the intravalley scattering with acoustic phonons, ionized impurities, acoustic piezoelectric phonons, polar optical phonons, and for the L-valleys also the scattering with optical nonpolar phonons; (ii) the intervalley scattering with the optical nonpolar phonons. Here all simulations refer to a GaAs bulk with a free electron concentration  $n = 10^{13} \text{ cm}^{-3}$ . With this value of impurity density and for the range of investigated temperatures ( $80 < T < 300 \text{ K}$ ) the Fermi temperature is much smaller than the electron temperature and degeneracy does not play any important role. We assume that all donors are ionized and that the free electron concentration is equal to the doping concentration. The complete set of n-type GaAs parameters used in the calculations is listed in Ref. (Persano Adorno et al., 2000). The scattering probabilities are calculated by the Fermi Golden Rule and assumed to be both field and spin independent; accordingly, the influence of the external fields is only indirect through the field-modified electron velocities. Nonlinear interactions of the field with the lattice and bound carriers are neglected. We also neglect electron-electron interactions and consider electrons to be independent (Kiselev & Kim, 2000). The spin polarization vector is included into the Monte Carlo algorithm and calculated for each free carrier, by taking into account the scattering-induced deviations of precession vector suffered after each collision.

The MC simulation is carried out by: (a) setting up the initial conditions of the system by giving to the electrons spin polarization, random momentum direction and kinetic energies with Maxwellian distribution depending on the lattice temperature; (b) determining the free flight time of each electron and updating the energy, the momentum and the spin polarization vector at the end of the free flight; (c) selecting a scattering process for each electron; (d) calculating the new value of energy (in case of inelastic scattering) and the scattering angles of each electron. The simulation continues from point (b). At fixed sampling time steps, small enough to properly update the particle motion, the average values of the physical quantities of interest, such as temperature, average electron velocity, spin polarization vector, are calculated.

### 3. Harmonic generation process in the presence of intense sub-terahertz electric fields

#### 3.1 A short introduction to the problem

The comprehension of the harmonic generation process in solid state nonlinear materials under the influence of far-infrared fields is important in perspective to use upconversion to create coherent sources of terahertz radiation and/or new devices for microwave and terahertz optics and electronics (Mikhailov, 2008; 2009). With this aim the process of harmonic emission arising from the interaction of semiconductor structures with intense radiation fields, having frequencies in the sub-THz range, has been both experimentally (Brazis et al., 1998; 2000; Moreau et al., 1999; Urban et al., 1995; 1996) and theoretically (Persano Adorno et al., 2000; 2001; 2004; 2007a;b; Shiktorov et al., 2002a;b; 2003a) widely investigated. Moreover, this field of research represents an useful tool for the general understanding of several features of the highly non linear processes of carrier transport in low-doped semiconductors. The basic

physical mechanism yielding harmonic generation in these materials, under cyclostationary conditions, is provided by carrier heating via scattering mechanisms. Indeed, the onset of a scattering mechanism usually results in a kink in the static velocity-field relation at some threshold electric field, corresponding to a characteristic energy (optical phonon energy, energy gap between the lower and upper valleys). Such kinks create a nonlinearity in the velocity-field relation which is responsible for velocity harmonic generation. A single alternating field generates only odd harmonics; absence of even harmonics is a consequence of the scattering cross section symmetry with respect to the inversion of the electron velocity direction and of the isotropy of the initial Maxwellian electron velocity distribution function. Both the addition of a static electric field or the mixing of two oscillating fields, having an integer ratio between the two frequencies, can break the inversion symmetry and generate also even harmonics, resulting, therefore, a sensitive means for the control of the emission rates of both even and odd harmonics. Indeed, it has been shown that an opportune manipulation of the relative intensity and polarization of the two fields, offers the possibility to produce coherent radiation beyond that achievable with a single field (Alekseev et al., 1999; Persano Adorno et al., 2003a;b; Romanov et al., 2004).

In the last decade, in order to better understand the harmonics emission process in atoms, gases and plasmas, in the presence of two laser fields, and to enhance its use in applications, several studies have been focused on the investigation of the harmonic polarization properties (Borca et al., 2000; Dudovich et al., 2006; Ferrante et al., 2000; 2004a;b; 2005; Song et al., 2003; Wang et al., 1999; Xia et al., 2007). In atoms, even for a linearly polarized driving laser, in the presence of an additional constant electric field, the harmonics are in general elliptically polarized (Borca et al., 2000). Moreover, the introduction of a second linearly polarized laser beam at low-intensity, collinear with the first one, can produce unusual polarization features of the generated harmonics (Borca et al., 2000). In plasmas driven by a laser field, in the presence of an additional electric static field, it has been found that the even polarization plane rotates with respect to that of the laser field and that the value of the rotation angle depends on the harmonic number (Ferrante et al., 2004a;b; 2005). This circumstance could allow to use harmonics also as a diagnostic tool (Mairesse et al., 2008).

Although the conversion efficiency in semiconductors subjected to two alternating electric fields has extensively been investigated, to the best of our knowledge, very little has been done in the study of the polarization of the emitted radiation. The primary focus of this part of the chapter is to show the effect of mixing two color fields, having an integer ratio between the two frequencies, on the polarization of both even and odd harmonics, that are generated in a low-doped GaAs bulk. We study the polarization of even and odd harmonics as a function of the angle between the direction of the two applied fields and compare the polarization obtained from the mixing of an oscillating field with a static electric field with that obtained in the presence of two periodic fields (Persano Adorno et al., 2007b; Persano Adorno, 2010).

### 3.2 Harmonic generation theory

The propagation of an electromagnetic wave along the  $z$  direction in a medium is described by the Maxwell equation

$$\frac{\partial^2 \vec{E}}{\partial z^2} - \frac{1}{c^2} \frac{\partial^2 \vec{E}}{\partial t^2} = \mu_0 \frac{\partial^2 \vec{P}}{\partial t^2} \quad (1)$$

where

$$\vec{P} = \epsilon_0(\chi_1 + \chi_2 E + \chi_3 E^2 + \dots) \vec{E} \quad (2)$$

is the polarization of the free electron gas in terms of the linear  $\chi_1$  and nonlinear  $\chi_2, \chi_3, \dots$  susceptibilities.

The source of the nonlinearity is the current density

$$\vec{j} = -ne \vec{v}(\vec{E}) \quad (3)$$

related to the polarization  $\vec{P}$  by

$$\vec{j} = \frac{\partial \vec{P}}{\partial t} \quad (4)$$

Expanding the electrons velocity and the electric field in terms of their Fourier components as

$$\vec{v} = \sum_q \vec{v}_q \exp\{-iq(\omega t - kz)\} \quad (5)$$

$$\vec{E} = \sum_q \vec{E}_q \exp\{-iq(\omega t - kz)\} \quad (6)$$

with  $\omega = 2\pi\nu$ , and taking only the leading term in the nonlinear part of the q-th component of the polarization  $\vec{P}$

$$\vec{P}_q^{NL} = \epsilon_0 \chi_q \vec{E}_1^q \quad (7)$$

we obtain a relation between the q-th component of the velocity and the susceptibility  $\chi_q$  given by

$$\chi_q = -\frac{inev_q}{q\omega E_1^q} \quad (8)$$

If we substitute the expansion for the electric field and use the above relation in the Maxwell equation, we obtain

$$\vec{E}_q[-q^2k^2 + \frac{1}{c^2}q^2\omega^2] = i\mu_0q\omega ne\vec{v}_q \quad (9)$$

By limiting the study of the harmonic generation efficiency to thin samples so as to reduce the loss, it is possible to not consider in the calculations the complex form of the dielectric function  $\epsilon(\nu)$  and neglect the field-dependent absorption. Assuming that the medium is transparent to the radiation at frequency  $\nu$ , i.e.  $\nu > \nu_p$ ,  $\nu_p$  being the donor plasma frequency, we can use the dispersion relation and calculate the efficiency of the harmonic generation at frequency  $\nu = q\nu_1$ , normalized to the fundamental one  $\nu_1$ , as (Persano Adorno et al., 2000; 2001):

$$\frac{I_\nu}{I_{\nu_1}} = \frac{1}{q^2} \frac{v_\nu^2}{v_{\nu_1}^2} \quad (10)$$

where  $v_\nu$  is the Fourier transform of the electron drift velocity, obtained via the three-dimensional multivalleys Monte Carlo simulation of the electron motion in the semiconductor. The spectra of emitted radiation are then reconstructed by the analysis of the velocity Fourier components.

### 3.3 Results and discussion

In all simulations the lattice temperature is  $T = 80$  K. The results are obtained in a stationary regime, after a transient time of a few picoseconds has elapsed. In all runs it is always present a linearly polarized oscillating electric field  $E_1(t)$ , having amplitude  $E_1 = 30$  kV/cm and frequency  $\nu_1 = 200$  GHz, which can rotate in the plane  $xy$ ; the effects on the harmonic generation process due to the introduction of a second field are studied. In order to investigate the polarization of the emitted radiation, as a function of the different geometries of the linear polarization of the two incident fields, the harmonic spectra along the  $x$ - and  $y$ -axis have been computed.

#### 3.3.1 Additional static field

We consider the GaAs bulk under the influence of the sum of the  $E_1(t)$  field with a constant electric field  $E_0 = \beta E_1$ , with  $\beta$  constant. The total field is described by the components:

$$E_x(t) = E_1 \cos(\varphi) \cos(2\pi\nu_1 t - k_1 z) + E_0 \quad (11)$$

$$E_y(t) = E_1 \sin(\varphi) \cos(2\pi\nu_1 t - k_1 z) \quad (12)$$

where  $\varphi$  is the angle between the static field  $E_0$ , directed along the  $x$ -axis, and the oscillating field  $E_1(t)$ , lying in the plane  $xy$ . The additional static electric field, by lowering the symmetry of the system, allows the generation of even harmonics, whose amplitudes increase with the strength of the static field (Persano Adorno et al., 2007b).

To choose the amplitude of the additional static field to be used to analyze the harmonic generation also for different geometries of the linear polarization of the two incident fields, a preliminary analysis of the efficiency of the generated harmonics, as a function of  $\beta$ , has been done.

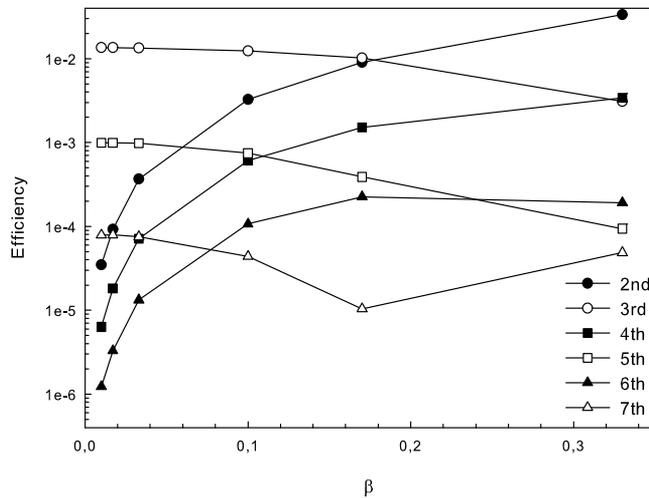


Fig. 1. Efficiency of the first harmonics as a function of  $\beta = E_0/E_1$ , calculated with  $E_1 = 30$  kV/cm and  $\nu_1 = 200$  GHz.

In Figure 1 we show the intensity of the first harmonics obtained with the total electric field directed along the  $x$ -axis ( $\varphi = 0^\circ$ ). For  $\beta \geq 0.1$  even and odd harmonics have comparable intensities; accordingly, to study the polarization of the emitted radiation, we adopt  $\beta = 0.1$ , i.e.  $E_0 = 3$  kV/cm, in all simulations.

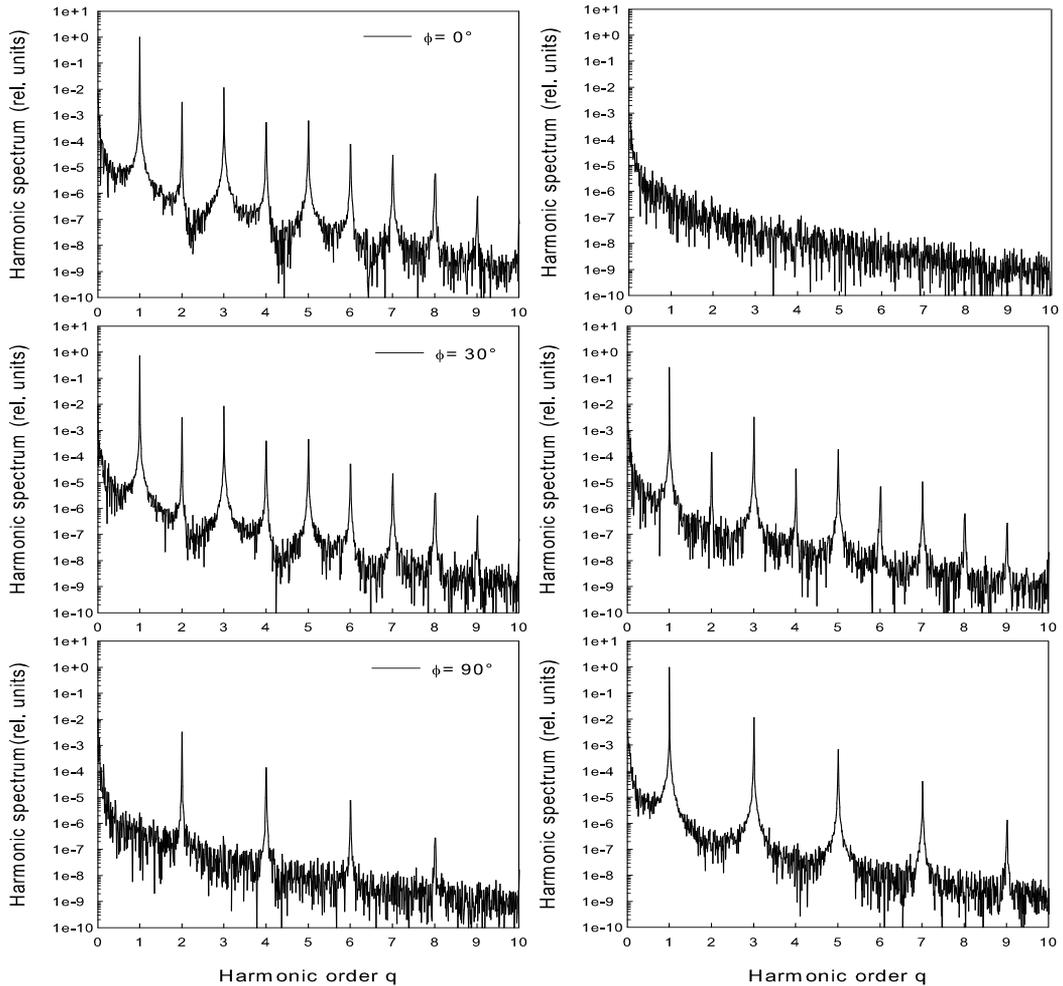


Fig. 2. Harmonic spectra along  $x$ -axis (left panels) and  $y$ -axis (right panels) at different  $\varphi$  as a function of the harmonic order  $q$ .  $E_1 = 30$  kV/cm,  $\nu_1 = 200$  GHz,  $\beta = 0.1$ .

Figure 2 shows the harmonic spectra along the  $x$  and the  $y$ -axis for different values of the angle  $\varphi$ . When  $\varphi = 0^\circ$ , along the  $x$ -axis we find, as expected, odd and even harmonics, and noise along the  $y$ -axis. Less obvious is the behaviour of the harmonic peaks when the two fields are not parallel, since the efficiency of both odd and even harmonics becomes function of the angle  $\varphi$ . In particular, when  $\varphi \neq 0^\circ$ , the spectrum along the  $y$ -axis contains non negligible even harmonics due to the presence of the static field, although along the  $y$  direction is present only  $E_1(t)$ . When  $\varphi = 90^\circ$  the spectrum along the  $x$ -axis shows only the even harmonics of the oscillating field, perpendicular to this direction, while along the  $y$  direction the spectrum contains only odd harmonics.

Figure 3 shows the angle  $\alpha$  between the  $x$ -axis and the direction of the first four even (upper panel) and odd harmonics (lower panel), as a function of the angle  $\varphi$ . Since the static field strength is small compared to the amplitude of  $E_1$ , odd harmonics are nearly polarized in the same direction of the sub-THz field. On the contrary, the coincidence of the polarization angle

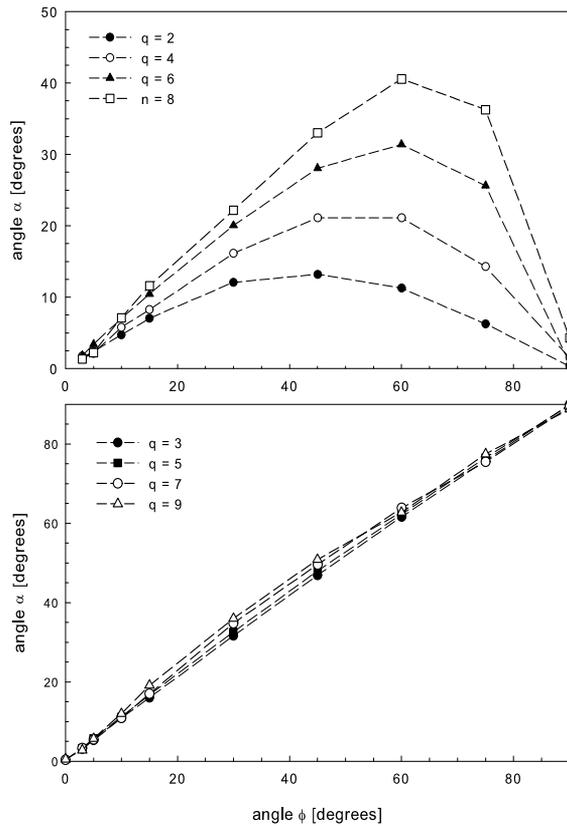


Fig. 3. Polarization angle  $\alpha$  of even (upper panel) and odd (lower panel) harmonics generated by the GaAs bulk as a function of the angle  $\phi$ .  $E_1 = 30$  kV/cm,  $\nu_1 = 200$  GHz,  $\beta = 0.1$ .

with that of the static field is not found for even harmonics. Moreover, even harmonics show a different polarization angle for each harmonic order. This behavior is close to that noted in a theoretical study of laser even harmonic generation by a plasma embedded in a static electric field (Ferrante et al., 2004a;b; 2005). Initially, for  $\phi < 20^\circ$ , even harmonics are polarized nearly in the same direction of  $E_1$ , then for  $20^\circ < \phi < 60^\circ$ , the angle  $\alpha$  increases more slowly than  $\phi$  and this effect appears more evident for lower values of the harmonic number  $q$ . Finally, for  $\phi > 60^\circ$  the polarization angle strongly decreases and, when  $\phi = 90^\circ$ , even harmonics are directed along the  $x$ -axis, as the constant field.

### 3.3.2 Additional oscillating field

The asymmetry effect induced by the presence of a static field may be also introduced by a second periodic field. In order to explore this case, we consider the GaAs bulk under the influence of a linearly polarized periodic electric field  $E(t)$  given by the sum of  $E_1(t)$  with a low-intensity field  $E_2(t)$ , having amplitude  $E_2 = 3$  kV/cm and frequency  $\nu_2 = 0.5 \nu_1$ , described by the components:

$$E_x(t) = E_1 \cos(\phi) \cos(2\pi\nu_1 t - k_1 z) + E_2 \cos(2\pi\nu_2 t - k_2 z) \quad (13)$$

$$E_y(t) = E_1 \sin(\phi) \cos(2\pi\nu_1 t - k_1 z) \quad (14)$$

where  $\varphi$  is the angle between the field  $E_2(t)$ , directed along the  $x$ -axis, and the field  $E_1(t)$ , lying in the plane  $xy$ .

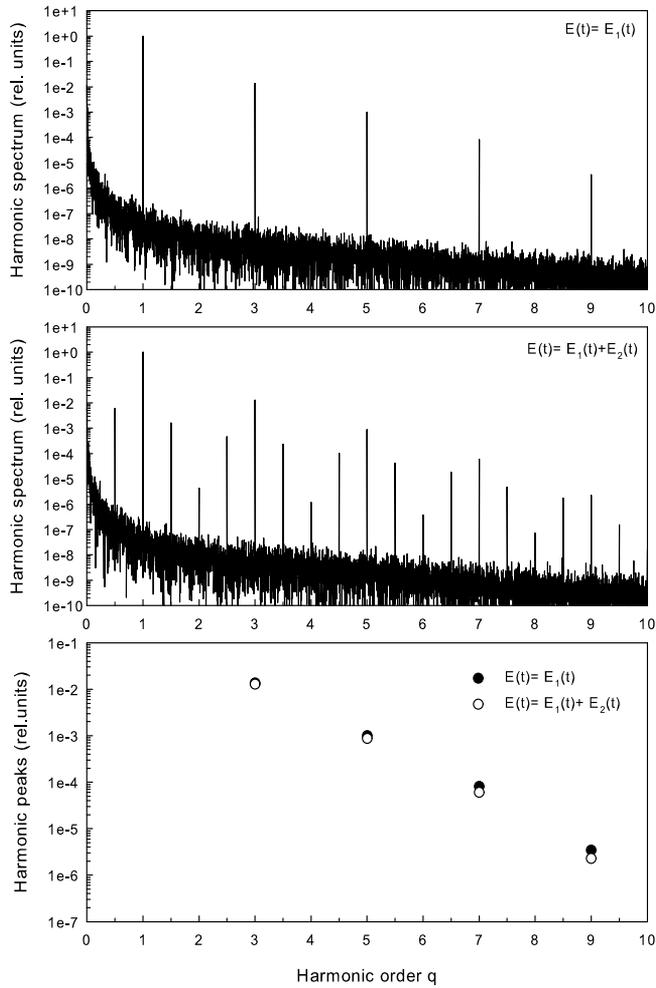


Fig. 4. Harmonic spectra obtained with the  $E_1$  field only (upper panel) and the  $E_2$  field parallel to the  $E_1$  field (central panel) as a function of the harmonic order  $q$ . Bottom panel show the intensity of the odd harmonic peaks in the two cases.  $E_1 = 30$  kV/cm,  $\nu_1=200$  GHz,  $E_2 = 3$  kV/cm,  $\nu_2=100$  GHz;  $\varphi = 0^\circ$ .

Harmonic spectra obtained with the  $E_1(t)$  field only (upper panel) and the  $E_2(t)$  field parallel to the  $E_1(t)$  field (central panel) are shown in Figure 4. When the two fields are parallel we found an enrichment of the spectrum with respect to the expected generation of odd harmonics of each field, because of the presence of satellite harmonics at mixed frequencies. In particular, as shown in the bottom panel of Figure 4, under field mixing conditions the peaks at the odd harmonics of the strong driving field frequency are present with amplitude nearly unchanged. Additionally, thanks to the nonlinear sum-frequency process, between pairs of odd harmonics of the frequency  $\nu_1$ , there are three peaks due to the presence of the field  $E_2(t)$ . This spectrum is very similar to that found in an experimental work on the generation of

harmonics in atomic gases (Perry & Krane, 1993). The intensity of these three additional peaks strongly depends on the relative angle between the two oscillating fields (see Figure 5).

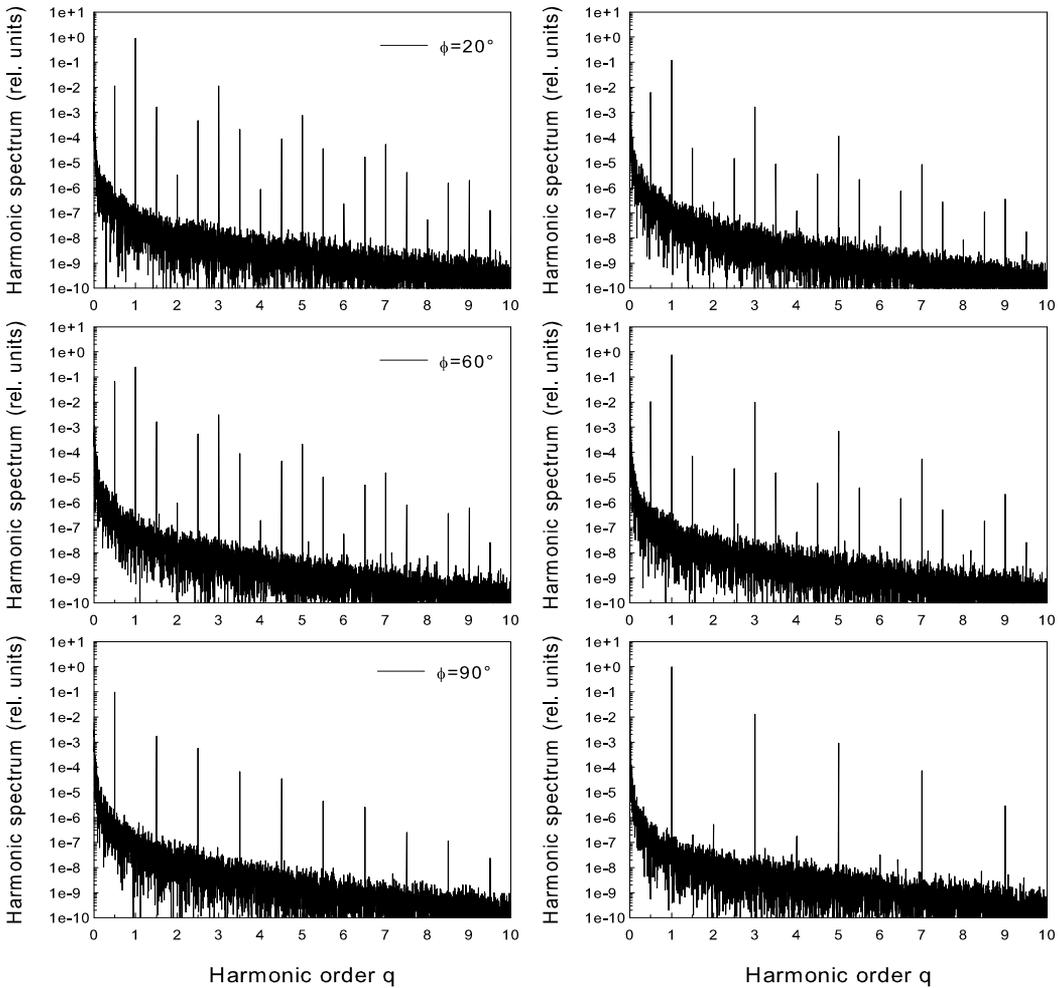


Fig. 5. Harmonic spectra along  $x$ -axis (left panels) and  $y$ -axis (right panels) at different  $\phi$  as a function of the harmonic order  $q$ .  $E_1 = 30$  kV/cm,  $\nu_1 = 200$  GHz,  $E_2 = 3$  kV/cm,  $\nu_2 = 100$  GHz.

When the field  $E_1(t)$  has a component different from zero along the  $y$ -axis, also the spectrum along this axis contains the three additional peaks due to the presence of the  $E_2(t)$  field, although along the  $y$  direction is present only the field  $E_1(t)$ . When the fields are orthogonal, i.e.  $\phi = 90^\circ$ , the spectrum along the  $y$ -axis still contains the even harmonics, while the peaks immediately on either side of the odd harmonics of the frequency  $\nu_1$  disappear; the spectrum along the  $x$ -axis contains only harmonics corresponding to odd multiples of the frequency  $\nu_2$ . As discussed in Ref. (Perry & Krane, 1993), the peak between the odd harmonics of the frequency  $\nu_1$ , which corresponds to an even harmonic of  $\nu_1$ , can be justified as true "even" harmonic, resulting from symmetry breaking of the Hamiltonian due to the presence of  $E_2(t)$ , or as resulting from different combinations of the two frequencies, as, for instance, from the sum of an odd harmonic of  $\nu_1$  plus the second harmonic of  $\nu_2$ . Also the peaks immediately on

each side of the odd harmonics of the frequency  $\nu_1$  can be attributed to true "odd" harmonics of  $E_2(t)$  or, for example, to the combination of an even harmonic of  $\nu_1$  plus or minus the signal at  $\nu_2$ . The analysis of the polarization of the emitted harmonics can help to highlight this aspect. Actually, the polarization of true "even" harmonics in two-colors experiment would be qualitatively similar to that predicted for an alternating field plus a static field, as long as the ratio of field strengths is comparable. Figure 6 shows the harmonic polarization plane for harmonic radiation corresponding to even (upper panel) and odd (central panel) multiples of the frequency  $\nu_1$ , as a function of the angle  $\varphi$  between the two oscillating fields. In the bottom panel is instead plotted the polarization angle of harmonics corresponding to odd multiples of the frequency  $\nu_2$ . Also in this case, "even" harmonics show different polarization plane for different harmonic order. Nevertheless, the dependence of the polarization angle  $\alpha$  on the relative direction of the two fields is different with respect to the case in which the even harmonics are due the introduction of a static electric field. In particular, in the presence of two periodic fields, for  $\varphi < 20^\circ$ , "even" order harmonics are polarized nearly in the same direction of the strong field  $E_1(t)$  and for  $20^\circ < \varphi < 60^\circ$  the angle  $\alpha$  increases more slowly than  $\varphi$ . Rather, for lower values of the harmonic number  $q$ ,  $\alpha$  remains almost constant. Instead, for  $\varphi > 60^\circ$  the polarization angle strongly increases and at  $\varphi = 90^\circ$ , "even" harmonics are directed along the  $y$ -axis. On the other hand, also the polarization plane of the "odd" harmonics of the frequency  $\nu_2$  does not coincide with the direction of the field  $E_2(t)$ , but it shows a dependence on the angle  $\varphi$ . Unexpectedly, the first "odd" harmonic is polarized up to  $\sim 40^\circ$  from the direction of  $E_2(t)$ . The polarization angle of the other three "odd" harmonics appears to be only slightly dependent on the angle  $\varphi$ , keeping it nearly constant at  $\sim 15^\circ$ , from the direction of  $E_2(t)$ .

### 3.4 Conclusions

By mixing a strong field with a weak field, oscillating at a frequency whose value is half of that of the high-intensity one, it is possible produce sum and difference-frequency radiation, including odd and even harmonics. The field-mixing does not affect the efficiency of the odd harmonics of the strong field and makes the spectra more rich. The behavior shown by the "even" harmonics shares some features with those occurring in the presence of an additional constant electric field, as, for example, the circumstance that even harmonics show different polarization plane for different harmonics order. However the dependence of the polarization angle of these "even" harmonics on the angle between the two incident fields, does not coincide with that obtained in the presence of an additional static electric field. On the contrary, for the odd harmonics of the high-intensity field there is no rotation of the polarization plane. Moreover, because the polarization angle of the "odd" harmonics of the low-frequency field does not coincide with its direction, we can conclude that these are not true "odd", but, as the "even" ones, are mainly due to the sum-frequency process.

## 4. Noise enhanced stability in semiconductor systems

### 4.1 A brief introduction to the problem

The presence of noise in experiments is generally considered a disturbance, especially studying the performance of semiconductor-based devices, where strong fluctuations can affect their response. Recently, however, an increasing interest has been directed towards the constructive aspects of noise on the dynamical response of non-linear systems. The effect of the interaction between an external source of fluctuations and an intrinsically noisy system has been analytically investigated for the first time by Vilar and Rubí in 2001. They have

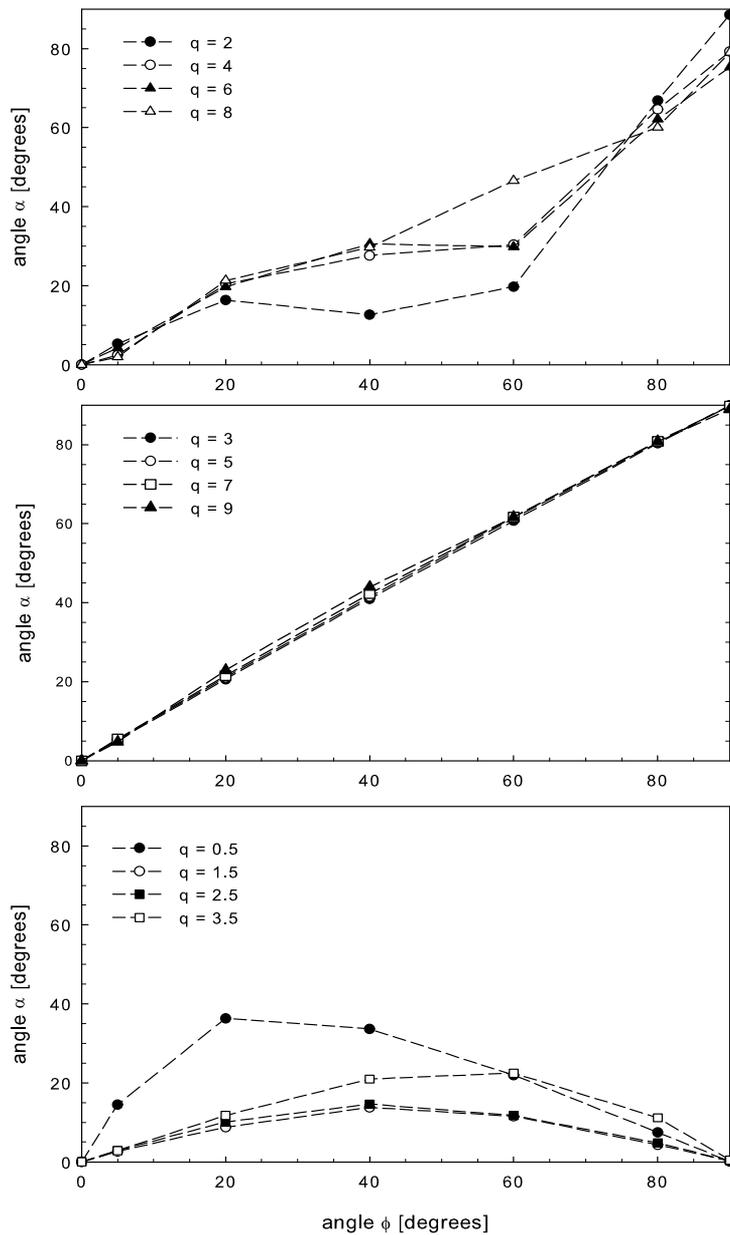


Fig. 6. Polarization angle  $\alpha$  of the generated harmonics as a function of the angle  $\phi$ . Upper and central panels show "even" and odd harmonics of the frequency  $\nu_1$ , respectively; the bottom panel refers to "odd" harmonics of the frequency  $\nu_2$ .  $E_1 = 30$  kV/cm,  $\nu_1 = 200$  GHz,  $E_2 = 3$  kV/cm,  $\nu_2 = 100$  GHz.

demonstrated that the spectral intensity of the output signal in a low frequency domain can be reduced by the addition of small amplitude noise on the input of the system (Vilar & Rubí, 2001).

In semiconductor bulk materials the possibility to reduce the diffusion noise by adding a correlated random contribution to a driving static electric field has been investigated by Varani and collaborators. Their numerical results, obtained by including energetic considerations in the theoretical analysis, have shown that, under specific conditions of the fluctuating electric field, it is possible to suppress the intrinsic noise in n-type GaAs semiconductors (Varani et al., 2005).

Recent studies of the electron velocity fluctuations in GaAs bulks driven by periodic electric fields, have shown that the spectral density strongly depends on the frequency of the applied field and critical modifications are observed when two mixed high-frequency large-amplitude periodic electric fields are used (Persano Adorno et al., 2008a). This means that the total power spectrum of the intrinsic noise is dependent on both the amplitude and the frequency of the excitation signals.

In this part of the chapter we focus our attention on the noise influence on the intrinsic carrier noise spectral density in low-doped n-type GaAs semiconductor driven by a high-frequency periodic electric field (Persano Adorno et al., 2008b; 2009a). The semiconductor intrinsic noise is obtained both by computing the velocity fluctuations correlation function and the spectral density (González et al., 2003; Shiktorov et al., 2003b) and directly calculating the variance of the electron velocity fluctuations. The effects caused by the addition of an external source of random perturbations are investigated by analyzing (i) the noise spectral density at the same frequency of the external driving field and (ii) the integrated spectral density (ISD), for different values of both the external noise amplitude and the noise correlation time. Numerical results show that, strictly depending on the correlation time, the presence of the external noise modifies the electron average velocity and significantly affects both the correlation function of its fluctuations and the internal noise spectrum of the system.

#### 4.2 Semiconductor noise calculation

The semiconductor bulk is driven by a fluctuating periodic electric field

$$E(t) = E_1 \cos(\omega t + \phi) + \eta(t) \quad (15)$$

with frequency  $\nu = \omega/2\pi$  and amplitude  $E_1$ . The random component of the electric field is modeled with an Ornstein-Uhlenbeck (OU) stochastic process  $\eta(t)$ , which obeys the following stochastic differential equation:

$$\frac{d\eta(t)}{dt} = -\frac{\eta(t)}{\tau_c} + \sqrt{\frac{2D}{\tau_c}} \zeta(t) \quad (16)$$

where  $\tau_c$  and  $D$  are, respectively, the correlation time and the variance of the OU process, and  $\zeta(t)$  is the Gaussian white noise with the autocorrelation  $\langle \zeta(t)\zeta(t') \rangle = \delta(t - t')$ . The OU correlation function is  $\langle \eta(t)\eta(t') \rangle = D \exp(-|t - t'|/\tau_c)$ . The changes on intrinsic noise properties are investigated by the statistical analysis of the autocorrelation function of the velocity fluctuations and of its mean spectral density. When the system is driven by a periodic electric field the correlation function  $C_{\delta v \delta v}(t, \tau)$  of the velocity fluctuations  $\delta v(t) = v(t) - \langle v(t) \rangle$  can be calculated as (González et al., 2003)

$$C_{\delta v \delta v}(t, \tau) = \left\langle v\left(t - \frac{\tau}{2}\right) v\left(t + \frac{\tau}{2}\right) \right\rangle - \left\langle v\left(t - \frac{\tau}{2}\right) \right\rangle \left\langle v\left(t + \frac{\tau}{2}\right) \right\rangle \quad (17)$$

in which  $\tau$  is the correlation time and the average is meant over a sequence of equivalent time moments  $t = s + mT$ , with  $s$  belonging to the time interval  $[0, T]$  ( $T$  is the field period) and  $m$  is an integer. This two-time symmetric correlation function eliminates any regular contribution and describes only the fluctuating part of  $v(t)$ . By averaging over the whole set of values of  $t$  within the period  $T$ , the velocity autocorrelation function becomes

$$C_{\delta v \delta v}(\tau) = \frac{1}{T} \int_0^T C_{\delta v \delta v}(t, \tau) dt \quad (18)$$

As due to the Wiener-Kintchine theorem, the spectral density can be calculated as the Fourier transform of  $C_{\delta v \delta v}(\tau)$ . In the computations of the autocorrelation function we have considered  $10^3$  possible initial values of  $s$  and a total number of equivalent time moments  $m \cong 10^6$ .

Intrinsic noise has been investigated also by estimating directly the electron velocity variance. This calculation has been performed separately for each energy valley, following the same method of equivalent time moments described above (Persano Adorno et al., 2009a).

#### 4.3 Physical model of intrinsic noise

When the semiconductor is driven by a static electric field, the shape of the spectral density of electron velocity fluctuations is exclusively determined by the strength of the applied field. For amplitudes smaller than the threshold field  $E_G$  (Gunn Field) for intervalley transitions, the diffusion is the most relevant source of noise, while, for  $E_0 > E_G$ , the complex structure of the semiconductor becomes relevant and random transitions of carriers among the available energy valleys must be taken into account. In this case, the intrinsic noise is mainly determined by a partition noise, caused by stochastic carrier transitions between regions characterized by different dynamical properties (intervalley transfers) in momentum space. The partition noise is characterized by a pronounced peak in the spectral density at a frequency  $\nu_G$ , which can be defined the "natural" transition frequency of the system between the valleys (Persano Adorno et al., 2008a).

Under cyclostationary conditions, the noise behaviour depends on both the amplitude and the frequency of the applied field. In particular, it is similar to that of the static field case only for very low-frequency fields ( $\nu \ll \nu_G$ ). On the contrary, for frequencies  $\nu \gtrsim \nu_G$ , the intervalley transfers are driven by the external field, the system enters in a forced regime of oscillations and the velocity fluctuations become time correlated (Persano Adorno et al., 2008a). In this case, the spectral density exhibits a peak centered around the frequency of the periodic signal and a significant enhancement in the low-frequency region.

#### 4.4 Numerical results and discussion

In order to neglect thermal noise contribution and to highlight the partition noise effects we have chosen as lattice temperature  $T = 80$  K. The spectral density of the electron velocity fluctuations has been studied by adopting a fluctuating periodic electric field with frequency  $\nu = 500$  GHz. The amplitude of this field has been chosen on the base of a preliminary analysis of both the variance of velocity fluctuations and the spectral density  $S_0(E)$  at zero frequency, as a function of the amplitude of the oscillating field. The most favorable condition to obtain a noise suppression effect in our system is reached when  $d^2 S_0(E)/dE^2$  is negative and the variance of velocity fluctuations exhibits a maximum (Varani et al., 2005; Vilar & Rubí, 2001). In accordance with the results shown in figures 1a and 1b of Ref. (Persano Adorno et al., 2009a), we have chosen a driving electric field with amplitude  $E_1 = 10$  kV/cm and frequency  $\nu = 500$  GHz.

In the absence of external noise, the amplitude of the forcing field is large enough to switch on intervalley transitions from the  $\Gamma$  valley to the  $L$  valleys and, since the frequency  $\nu$  is of the same order than  $\nu_G$ , the electron velocity fluctuations are mainly determined by partition noise (Shiktorov et al., 2003b). Hence, the spectrum is characterized by the features described in section 4.3

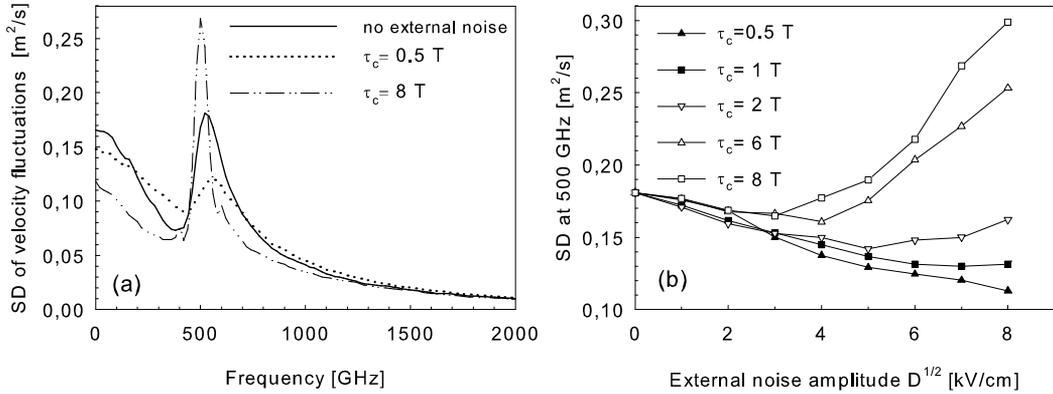


Fig. 7. (a) Spectral density (SD) of electron velocity fluctuations as a function of the frequency. Solid line is obtained in the absence of external noise; dotted line describes the results obtained with  $D^{1/2} = 7 \text{ kV}/\text{cm}$  and  $\tau_c = 1 \text{ ps} = 0.5 \text{ T}$ ; dashed-dotted line is obtained with  $D^{1/2} = 7 \text{ kV}/\text{cm}$  and  $\tau_c = 16 \text{ ps} = 8 \text{ T}$ . (b) Height of the peak in the SD of electron velocity fluctuations as a function of the external noise amplitude for different values of the correlation time  $\tau_c$  of the external noise source.

In figure 7a we show how the spectral density of electron velocity fluctuations is modified by the presence of noise. The addition of an external source of fluctuations to the driving electric field strongly changes the spectrum and, in particular, the height of the peak around 500 GHz, in a way that critically depends on the OU correlation time. In figure 7b we plot the maximum of the spectral density at the frequency of the driving field as a function of the external noise amplitude  $D^{1/2}$ , for five different values of  $\tau_c$ . An interesting nonlinear behavior of this quantity is observed for increasing noise intensities and correlation times. In particular, for values of  $\tau_c$  smaller than or equal to the period  $T$  of the oscillating electric field, the spectral density at 500 GHz shows a monotonic decreasing trend with increasing noise amplitude. For values of  $\tau_c$  greater than  $T$ , the spectral density is reduced only for small amplitudes of the external noise, while an enhancement of the peak is observed for greater intensities. When the intrinsic noise is mainly due to the partition effect, the height of the peak in the spectral density depends on the population of the different valleys and it reaches a maximum when the populations are nearly at the same level (Nougier, 1994; Shiktorov et al., 2003b). Since the "effective" electric field experienced by electrons in the presence of a fluctuating field is different, the number of intervalley transitions changes with respect to the case in which the external source of noise is absent. This fact can be responsible of the observed changes on the peak of the spectral density.

The dependence of the intrinsic noise suppression effect on the amplitude and the correlation time of the external source of fluctuations has been investigated also by studying the integrated spectral density (ISD), i. e. the total noise power, as a function of the OU noise amplitude, for three different values of  $\tau_c$ . In figure 8 we show a clear reduction of the ISD in the presence of added noise. In particular, for each value of the correlation time, we find a

range of  $D^{1/2}$  in which the electric field fluctuations reduce the semiconductor intrinsic noise. This effect is more evident for higher correlation times.

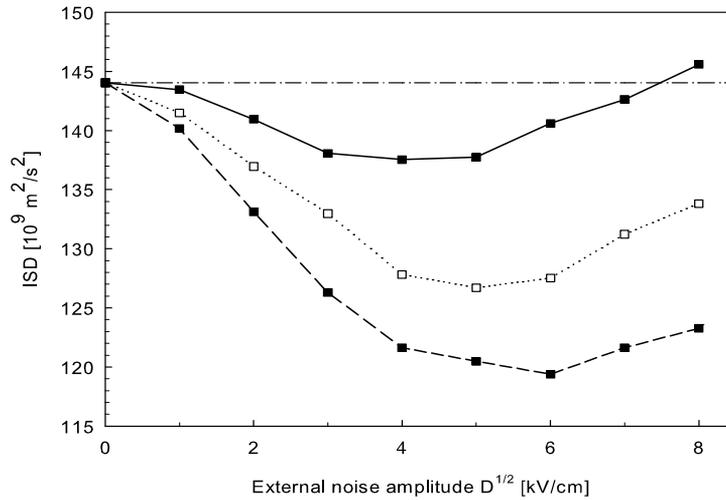


Fig. 8. Integrated spectral density of electron velocity fluctuations as a function of the external noise amplitude. Solid line:  $\tau_c = 0.5 T$ ; dotted line:  $\tau_c = 2 T$ ; dashed line:  $\tau_c = 8 T$ . Dashed-dotted line is the semiconductor intrinsic noise level

From a microscopic point of view, this suppression can arise from the fact that the fluctuating electric field forces the carriers to visit regions of the momentum space characterized by a smaller variance with respect to the case of zero noise (Walton & Visscher, 2004). We have investigated the details of the electron dynamics under the fluctuating electric field by analyzing for different correlation times the relative occupation time and the velocity variance separately in different valleys. In figure 9 (right panels) we show that, when the noise amplitude increases, the electron occupation time of the  $\Gamma$  valley decreases and the corresponding times calculated for the  $L$  and  $X$  valleys increase. This behavior is expected because the addition of fluctuations to the driving electric field leads to an increase of scattering events which are responsible for an increase of the transitions from the  $\Gamma$  valley to valleys at higher energy. This behavior depends on the correlation time of the external noise source. In particular, for a fixed value of the external noise intensity, the effect of reduction of the relative occupation time for the  $\Gamma$  valley and the corresponding increase for the  $L$  valleys is more pronounced for shorter correlation times.

Less obvious is the behavior of the electron velocity variance evidenced in figure 9 (left panels). In fact, while the common experience would suggest an increase of the velocity variance when the external noise amplitude grows up, we find that, depending on the value of the correlation time, the velocity variance in the  $\Gamma$  valley can be reduced in a specific range of the noise amplitude. An increasing trend is instead observed for the  $L$  and  $X$  valleys. The reduction of the electron velocity variance observed in the  $\Gamma$  valley for  $\tau_c = 2 T$  and  $D^{1/2}$  between 1 and 6 and, even more, for  $\tau_c = 8 T$  and  $D^{1/2}$  between 1 and 8, represents an intrinsic effect of the dynamics of electrons in the  $\Gamma$  valley without taking into consideration any transfer to valleys characterized by different dynamical properties. This effect of noise-induced stability can explain the longer residence times of electrons in the  $\Gamma$ -valley at higher correlation times.

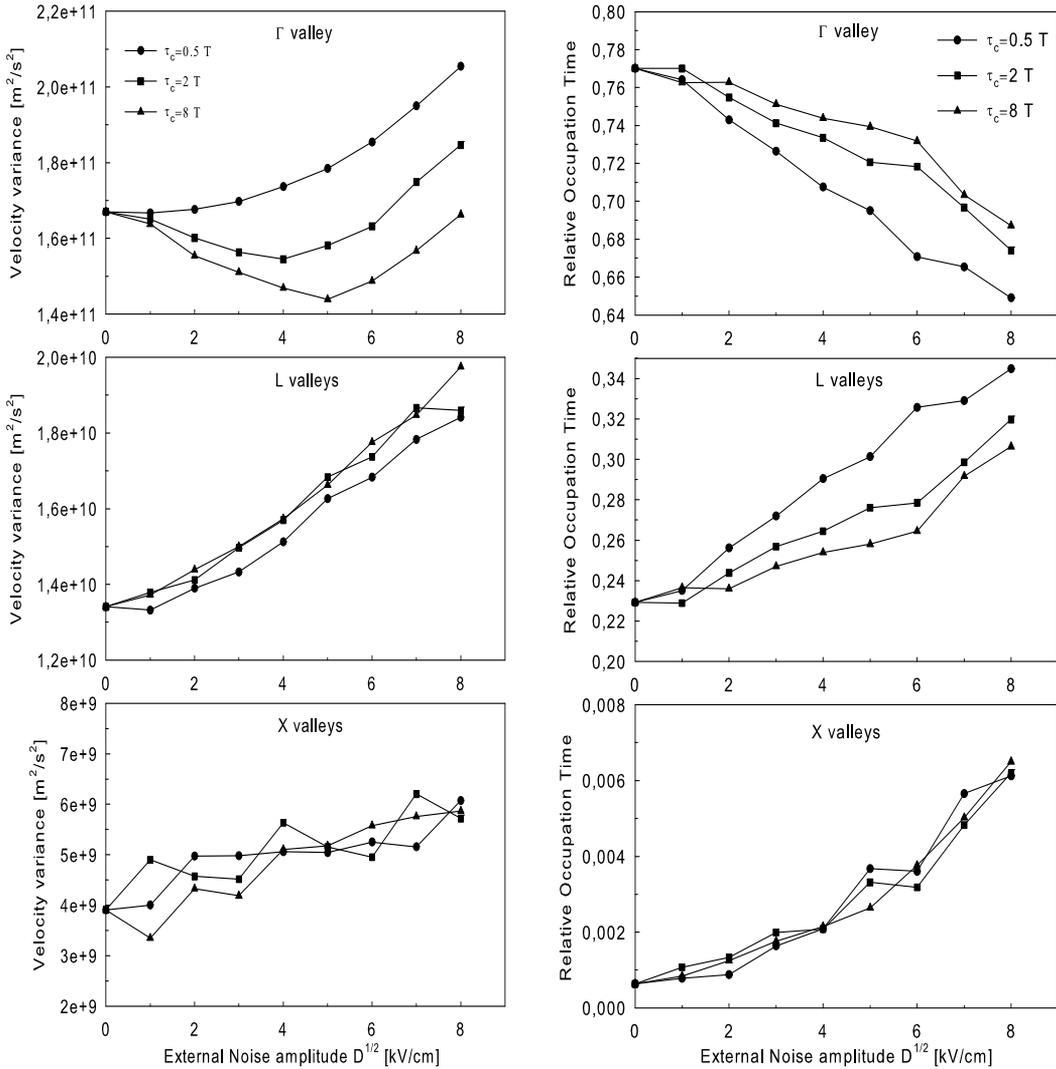


Fig. 9. Variance of the electron velocity fluctuations (left panels) and relative occupation time (right panels) as a function of the external noise amplitude, for three different correlation times.

#### 4.5 Conclusions

A less noisy response in the presence of a driving periodic electric field containing time-correlated fluctuations is observed. Both the amplitude and the correlation time of the electric field fluctuations are crucial parameters for the intrinsic noise reduction effect. Previous studies ascribe the reduction of the electron velocity fluctuations to an overall effect of intervalley transfers. Our study on the electron velocity variance, calculated separately for every single energy valley of the semiconductor, shows that the velocity variance of an electron moving in the  $\Gamma$ -valley is reduced by the presence of correlated noise, independently from the transitions to upper valleys, bringing to longer residence times. This effect of noise enhanced stability (NES) arises from the fact that the transport dynamics of electrons in the

semiconductor receives a benefit by the constructive interplay between the fluctuating electric field and the intrinsic noise of the system.

Moreover, for a fixed value of the external noise amplitude, a very unexpected non-monotonic behavior of the integrated spectral density as a function of the noise correlation time, in a wider range of  $\tau_c$ , is discussed in (Persano Adorno et al., 2009b).

## 5. Influence of transport conditions on the electron spin depolarization

### 5.1 Preliminary remarks

For an extensive utilization of spintronic devices, the features of spin decoherence at relatively high temperature, jointly with the influence of transport conditions, should be fully understood. In last decade there has been a lot of experimental works in which the influence of transport conditions on relaxation of spins in semiconductors has been investigated (Beck et al., 2006; Dzhioev et al., 2004; Furis et al., 2006; Hägele et al., 1998; Hruška et al., 2006; Kikkawa & Awschalom, 1998; Sanada et al., 2002). Even though for high speed transfer of information, high external electric fields must be used, up now only the influence of low electric fields ( $F < 0.1$  kV/cm) on coherent spin transport has been investigated and very little is known about the effects of higher electric fields (Sanada et al., 2002) or high lattice temperatures. Very recently, electrical injection of spin polarization in n-type and p-type silicon at room-temperature have been experimentally carried out (Dash et al., 2009). These promising experimental results for development of spintronic devices suggest that it is important investigate the spin coherence up to room temperature.

The temporal evolution of the spin and the evolution of the momentum of an electron cannot be separated. The spin depolarization rates are functionals of the electron distribution function in momentum space which continuously evolves with time when an electric field is applied to drive the transport. Thus, the dephasing rate is a dynamic variable that needs to be treated self-consistently in step with the dynamic evolution of the electron's momentum. A way to solve this problem is to describe the transport of spin polarization by making use of Boltzmann-like kinetic equations. This can be done within the density matrix approach (Ivchenko et al., 1990), methods of nonequilibrium Green's functions, as the microscopic kinetic spin Bloch equation approach (Jiang & Wu, 2009; Weng & Wu, 2003; Weng et al., 2004; Wu & Ning, 2000; Wu et al., 2010; Zhang et al., 2008), or Wigner functions (Mishchenko & Halperin, 2003; Saikin, 2004), where spin property is accounted for starting from quantum mechanics equations. Another way is to use a Monte Carlo approach, by taking into account the spin polarization dynamics with the inclusion in the code of the precession mechanism of the spin polarization vector (Barry et al., 2003; Bournel et al., 2000; Kiselev & Kim, 2000; Pershin, 2005; Pramanik et al., 2003; Saikin et al., 2003; 2006; Shen et al., 2004; Spezia et al., 2010; Spezia et al., , in press). Both methods allow to include the relevant spin relaxation phenomena for electron systems and take into account the details of electron scattering mechanisms, material properties and specific device design; their predictions have been found to be in good agreement with experiments.

Earlier Monte Carlo simulation has revealed that the presence of an external electric field can accentuate spin relaxation in GaAs bulk materials (Barry et al., 2003). However, a comprehensive theoretical investigation of the influence of transport conditions on the spin depolarization in semiconductor bulk structures, in a wide range of values of temperature, doping density and amplitude of external fields, is still lacking. In this last part of the chapter, solving the transport and spin dynamics stochastic differential equations by a semiclassical Monte Carlo approach, the spin lifetimes and depolarization lengths of an

ensemble of electrons, for intermediate values of the electric field (0.1 – 2.5 kV/cm) and lattice temperatures in the range  $10 < T < 300$  K, are estimated (Spezia et al., 2010). A detailed analysis of the doping density effect on the fast process of spin depolarization of drifting electrons in GaAs bulks, below the metal-to-insulator transition, can be found in Spezia et al., (in press).

## 5.2 Spin relaxation dynamics

Spin dephasing may be caused by interactions with local magnetic fields originating from nuclei and spin-orbit interactions or magnetic impurities. The most relevant spin relaxation mechanisms for an electron system under non degenerate regime are: (i) the Elliott-Yafet (EY) mechanism, in which electron spins have a small chance to flip during each scattering, due to the spin mixing in the conduction band (Elliott, 1954; Yafet, 1963); (ii) the Dyakonov-Perel (DP) mechanism, based on the spin-orbit splitting of the conduction band in non-centrosymmetric semiconductors, in which the electron spins decay due to their precession around the  $\mathbf{k}$ -dependent spin-orbit fields (inhomogeneous broadening) during the free flight between two successive scattering events (Dyakonov & Perel, 1971; Dyakonov, 2006); (iii) the Bir-Aronov-Pikus (BAP) mechanism, in which electrons exchange their spins with holes (Bir et al., 1976). Hyperfine interaction is another mechanism, usually important for spin relaxation of localized electrons, but ineffective in metallic regime where most of the carriers are in extended states (Abragam, 1961; Paget et al., 1977; Pikus & Titkov, 1989).

Previous theoretical (Jiang & Wu, 2009; Wu et al., 2010) and experimental (Litvinenko et al., 2010) investigations indicate that the the EY mechanism is totally irrelevant on electron spin relaxation in n-type III-V semiconductors. For this reason we analyze the spin depolarization of drifting electrons in n-type GaAs semiconductors by considering only the D'yakonov-Perel process.

By following the semiclassical formalism, the term of the single-electron Hamiltonian which accounts for the spin-orbit interaction can be written as

$$H_{SO} = \frac{\hbar}{2} \vec{\sigma} \cdot \vec{\Omega}. \quad (19)$$

It represents the energy of electron spins precessing around an effective magnetic field [ $\vec{B} = \hbar\vec{\Omega}/\mu_B g$ ] with angular frequency  $\vec{\Omega}$ , which depends on the orientation of the electron momentum vector with respect to the crystal axes. Near the bottom of the  $\Gamma$ -valley, the precession vector can be written as (Pikus & Titkov, 1989)

$$\vec{\Omega}_\Gamma = \beta_\Gamma [k_x(k_y^2 - k_z^2)\hat{x} + k_y(k_z^2 - k_x^2)\hat{y} + k_z(k_x^2 - k_y^2)\hat{z}] \quad (20)$$

In equation (20),  $k_i$  ( $i = x, y, z$ ) are the components of the electron wave vector.  $\beta_\Gamma$  is the spin-orbit coupling coefficient, a crucial parameter for the simulation of spin polarization. In  $\Gamma$ -valley we consider the effects of nonparabolicity on the spin-orbit splitting by using (Pikus & Titkov, 1989),

$$\beta_\Gamma = \frac{\alpha \hbar^2}{m \sqrt{2mE_g}} \left( 1 - \frac{E(\vec{k})}{E_g} \frac{9 - 7\eta + 2\eta^2}{3 - \eta} \right) \quad (21)$$

where  $\alpha = 0.029$  is a dimensionless material-specific parameter,  $\eta = \Delta/(E_g + \Delta)$ , with  $\Delta = 0.341$  eV the spin-orbit splitting of the valence band,  $E_g$  is the energy separation between the conduction band and valence band at the  $\Gamma$  point,  $m$  the effective mass and  $E(\vec{k})$  the electron energy.

The quantum-mechanical description of electron spin evolution is equivalent to that of the classical momentum  $\vec{S}$  experiencing the effective magnetic field, as described by the equation of motion

$$\frac{d\vec{S}}{dt} = \vec{\Omega} \times \vec{S}. \quad (22)$$

Every scattering event changes the orientation of the effective magnetic field  $\vec{B}$  (that strongly depends on  $\vec{k}$ ) and the direction of the spin precession axis.

### 5.3 Calculation of spin depolarization times and lengths

All simulations are performed by using a temporal step of 10 fs and an ensemble of  $5 \cdot 10^4$  electrons to collect spin statistics. The initial non-equilibrium spin polarization decays with time as the electrons, driven by a static electric field, move through the medium, experiencing elastic and anelastic collisions. Since scattering events randomize the direction of  $\vec{\Omega}$ , during the motion, the polarization vector of the electron spin experiences a slow angular diffusion. The dephasing of each individual electron spin produces a distribution of spin states that results in an effective depolarization, which is calculated by ensemble-averaging over the spin of all the electrons.

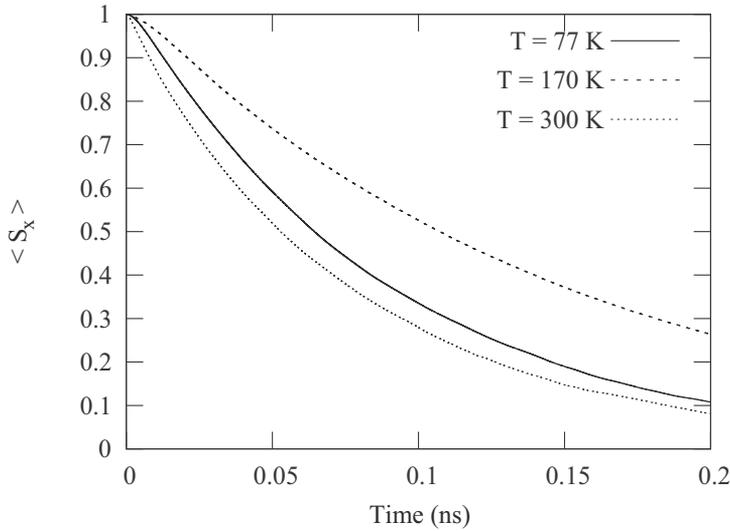


Fig. 10. Spin polarization  $\langle S_x \rangle$  as a function of time for three different values of the lattice temperature  $T$ ,  $E_0 = 0.1$  kV/cm.

The simulation of the spin relaxation starts with all the electrons of the ensemble initially polarized ( $\langle \vec{S} \rangle = 1$ ) along the  $x$ -axis at the injection plane ( $x_0 = 0$ ). After a transient time of typically  $10^4$  time steps, long enough to achieve the steady-state transport regime, the electron spins are initialized, the spin relaxation begins and the quantity  $\langle \vec{S} \rangle$  is calculated as a function of time. In Fig. 10, we show the electron average polarization  $\langle S_x \rangle$ , calculated as a function of time in the presence of an electric field, having amplitude  $E_0 = 0.1$  kV/cm, directed along the  $x$ -axis, for three different values of temperature. In order to extract the characteristic time  $\tau$  of the spin relaxation, the obtained trend of the spin dephasing is fitted by the following

exponentially time decaying law

$$\langle S_x \rangle(t) = A \cdot \exp(-t/\tau), \quad (23)$$

with  $A$  a normalization factor. The depolarization length  $L$  is the corresponding value of the distance traveled by the center of mass of the electron cloud from the injection plane.

#### 5.4 Numerical results and discussion

In Fig. 11 we plot the spin depolarization length  $L$  [panel (a)] and the spin depolarization time  $\tau$  [panel (b)] as a function of the electric field amplitude, for three values of the lattice temperature. The spin relaxation lengths show a marked maximum that rapidly reduces its intensity, widens and moves towards higher electric field amplitudes with the increasing of the temperature. For temperatures  $T \leq 150$  K the decoherence times plotted in Fig. 11 (b) show a non-monotonic behavior. For  $E_0 > 0.5$  kV/cm,  $\tau$  lightly depends on the temperature up to  $T \sim 150$  K. At higher temperatures, the spin electron relaxation time becomes a monotonic decreasing function of the electric field intensity. The presence of maxima in the spin depolarization length at intermediate fields can be explained by the interplay between two competing factors: in the linear regime, as the field enlarges, the electron momentum and the drift velocity increase in the direction of the field. On the other hand, the increased electron momentum also brings about a stronger effective magnetic field, as shown in Eq. 20 (Barry et al., 2003). Consequently, the electron precession frequency becomes higher, resulting in faster spin relaxation (i.e., shorter spin relaxation time). For  $E_0 < 0.5$  kV/cm and  $T \leq 150$  K the non-monotonic behavior of the relaxation time reflects the complex scenario described above, caused by the triggering of scattering mechanisms having different rates of occurrence.

In Fig. 12 we show the spin electron relaxation length  $L$  [panel (a)] and the spin depolarization time  $\tau$  [panel (b)] as a function of the lattice temperature, for five values of the electric field amplitude. For a fixed electric field,  $L$  is a monotonic decreasing function of the temperature. When  $E_0 = 0.5$  kV/cm,  $L$  shows its maximum value, remaining greater than  $35 \mu\text{m}$  up to  $T \simeq 80$  K. Furthermore, for field amplitudes greater than 1 kV/cm, the spin depolarization length remains almost constant for  $T < 100$  K. At room temperature the maximum value of  $L$  ( $\sim 6 \mu\text{m}$ ) is obtained for  $E_0 \geq 1$  kV/cm. The relaxation time  $\tau$  shows, instead, a non-monotonic behavior with the temperature [see Fig. 12 (b)]. In particular, the curve obtained with  $E_0 = 0.1$  kV/cm exhibits a minimum at  $T \sim 80$  K and an increase in the range 80 – 150 K. For temperatures greater than 150 K, all curves with a field strength up to 0.5 kV/cm show a common decreasing trend. The longest value of spin coherence time is achieved for the field amplitude  $E_0 = 0.5$  kV/cm for the entire range of temperatures. For higher values of  $E_0$ , the spin depolarization time strongly decreases, becoming nearly temperature-independent for  $E_0 > 1.5$  kV/cm. As the temperature increases, the scattering rate increases too, and hence the ensemble of spins loses its spatial order faster, resulting in a faster spin relaxation. This temperature dependence becomes less evident at higher amplitudes of the driving electric field, where, because of the greater drift velocities, the polarization loss is mainly due to the strong effective magnetic field. At very low electric fields, the spin dephasing is, instead, primarily caused by the multiple scattering events. The nonmonotonicity of  $\tau$  can be ascribed by the progressive change, with the increase of the temperature, of the dominant collisional mechanism from acoustic phonons and ionized impurities to polar optical phonons (Dzhioev et al., 2004). Following the theory of D'yakonov-Perel,  $\tau^{-1}$  is proportional to the third power of the temperature  $T$  and linearly depends on the momentum relaxation time  $\tau_p$  (Dyakonov & Perel, 1971). An increase of the temperature initially leads to a slightly decrease of  $\tau$ ; for

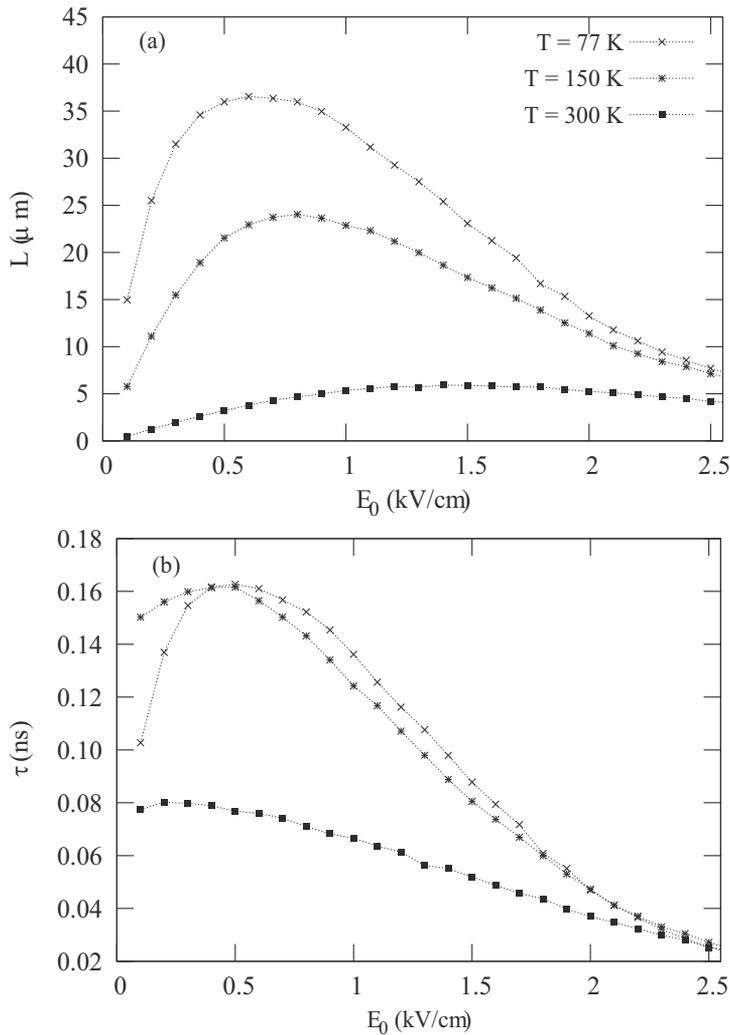


Fig. 11. Spin depolarization length  $L$  (a) and spin depolarization time  $\tau$  (b) as a function of the electric field amplitude  $E_0$ , for three values of the lattice temperature  $T$ .

temperatures greater than  $\sim 100\text{ K}$  the electrons start to experience scattering by polar optical phonons. This switching on leads to an abrupt decrease of  $\tau_p$  that, for lattice temperatures in the range  $100 - 150\text{ K}$ , results more effective than the increase of  $T$ , giving rise to the observed increase of  $\tau$ . For temperatures greater than  $150\text{ K}$  this latter effect is no more relevant.

### 5.5 Conclusions

We have estimated the spin mean lifetimes and depolarization lengths of an ensemble of conduction electrons in lightly doped n-type GaAs crystals, in a wide range of both lattice temperatures ( $10 < T < 300\text{ K}$ ) and field intensities ( $0.1 < E_0 < 2.5\text{ kV/cm}$ ), finding that, under particular conditions, also at temperatures greater than the liquid-helium temperature, it is possible to obtain very long spin relaxation times and relaxation lengths. These are

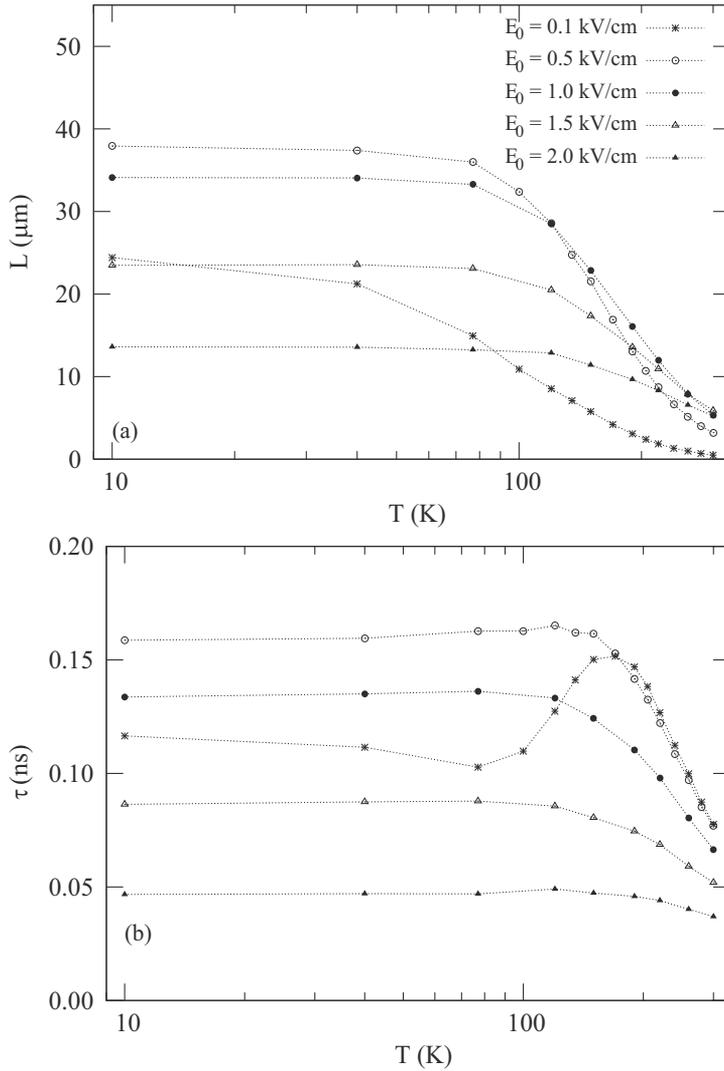


Fig. 12. Spin depolarization length  $L$ (a) and spin depolarization time  $\tau$  (b) as a function of the temperature  $T$ , for five values of the electric field amplitude  $E_0$ .

essential for the high performance of spin-based devices, in order to extend the functionality of conventional devices to higher working temperatures and higher electric field amplitudes and to allow the development of new information processing systems. In particular, for  $E_0 = 0.5$  kV/cm we achieve the longer value of spin lifetime ( $\tau > 0.15$  ns) up to a temperature  $T = 150$  K. At room temperatures, we obtain a coherence length of about  $6 \mu\text{m}$ , nearly independent from the intensity of the electric field. Furthermore, depending on the interplay between the external electric field and the different collisional mechanisms with increasing electron energy, we find very interesting nonmonotonic behavior of spin lifetimes and depolarization lengths as a function of temperature and electric field amplitude.

## 6. Acknowledgements

This work was partially supported by MIUR and CNISM-INFN. DPA would like to thank Prof. G. Ferrante, Prof. M. Zarcone, Dr. M.C. Capizzo (Harmonic generation processes), Prof. B. Spagnolo, Dr. N. Pizzolato and Dr. S. Spezia (hot-carrier noise and spin relaxation) for their interesting ideas, helpful discussions and excellent collaborations and, in particular, N. Pizzolato and S. Spezia for the proofreading of the whole manuscript. Moreover, the author is very thankful to Prof. M.W. Wu, who improved the spintronics part with valuable comments and suggestions.

## 7. References

- Abraham, A. (1961) *The Principles of Nuclear Magnetism*, Clarendon Press, Oxford.
- Alekseev, K.N.; Erementchouk, M.V. & Kusmartsev, F.V. (1999). Direct current generation due to wave mixing in semiconductors. *Europhys. Lett.*, 47, 595.
- Barry, E.A.; Kiselev A.A. & Kim K.W. (2003). Electron spin relaxation under drift in GaAs. *Appl. Phys. Lett.*, 82, 3686.
- Beck, M.; Metzner, C.; Malzer, S. & Döhler G.H. (2006). Spin lifetimes and strain-controlled spin precession of drifting electrons in GaAs. *Europhys. Lett.*, 75, 597.
- Bir, G.L.; Aronov, A.G. & Pikus, G.E. (1976). Spin relaxation of electrons due to scattering by holes. *Sov. Phys. - JETP* 42, 705
- Borca, B.; Flegel, A.V.; Frolov, M.V.; Manakov, N.L.; Milosevic, D.B. & Starace, A.F. (2000). Static-Electric-Field-Induced Polarization Effects in Harmonic Generation. *Phys. Rev. Lett.*, 85, 732.
- Bournel, A.; Dollfus, P.; Cassan, E. & Hesto P. (2000). Monte Carlo study of spin relaxation in AlGaAs/GaAs quantum wells. *Appl. Phys. Lett.*, 77, 2346.
- Brazis, R.; Raguotis, R. & Siegrist, M.R. (1998). Suitability of drift nonlinearity in Si, GaAs, and InP for high-power frequency converters with a 1 THz radiation output. *J. Appl. Phys.*, 84, 3474.
- Brazis, R.; Raguotis, R.; Moreau, Ph. & Siegrist, M.R. (2000). Enhanced Third-Order Nonlinearity in Semiconductors Giving Rise to 1 THz Radiation. *Int. J. Infrared Millim. Waves*, 21, 593.
- Dash, S.P.; Sharma, S.; Patel, R.S.; De Jong, M.P. & Jansen, R. (2009). Electrical creation of spin polarization in silicon at room temperature. *Nature* 462, 491.
- Dyakonov, M.I. & Perel V.I. (1971). Spin relaxation of conduction electrons in noncentrosymmetric semiconductors. *Sov. Phys. - Solid State*, 13, 3023.
- Dyakonov, M.I. (2006). Introduction to spin physics in semiconductors. *Physica E*, 35, 246.
- Dzhioev, R.I.; Kavokin, K.V.; Korenev, V.L.; Lazarev, M.V.; Poletaev, N.K.; Zakharchenya, B.P.; Stinaff, E.A.; Gammon, D.; Bracker, A.S. & Ware, M.E. (2004). Suppression of Dyakonov-Perel Spin Relaxation in High-Mobility n-GaAs. *Phys. Rev. Lett.*, 93, 216402.
- Dudovich, N.; Smirnova, O.; Levesque, J.; Mairesse, Y.; Ivanov, M.Yu.; Villeneuve, D.M. & Corkum, P.B. (2006). Measuring and controlling the birth of attosecond XUV pulses. *Nature Physics*, 2, 781.
- Elliott, R.J. (1954). Theory of the Effect of Spin-Orbit Coupling on Magnetic Resonance in Some Semiconductors. *Phys. Rev.*, 96, 266.

- Ferrante, G.; Zarccone, M. & Uryupin, S.A. (2000). Harmonic generation and wave mixing in a plasma in the presence of two linearly polarized laser fields. *J. Opt. Soc. Am. B*, 17, 1383.
- Ferrante, G.; Zarccone, M. & Uryupin, S.A. (2004a). Plasma radiation spectra in the presence of static electric and high-frequency radiation fields. *Eur. Phys. J. D*, 31, 77.
- Ferrante, G.; Zarccone, M. & Uryupin, S.A. (2004b). Laser even harmonics generation by a plasma embedded in a static electric field. *Laser Phys. Lett.*, 1, 167.
- Ferrante, G.; Zarccone, M. Uryupin, S.A. (2005). Even harmonics generation of high frequency radiation in current-carrying plasmas. *Physics of Plasmas*, 12, 052111.
- Furis, M.; Smith, D.L.; Crooker, S.A. & Reno, J.L. (2006). Bias-dependent electron spin lifetimes in n-GaAs and the role of donor impact ionization. *Appl. Phys. Lett.*, 89, 102102.
- González, T.; Pérez, S.; Starikov, E.; Shiktorov, P.; Gruzinskis, V.; Reggiani, L.; Varani, L. & Vaissiére, J.C. (2003). Microscopic investigation of large-signal noise in semiconductor materials and devices. *Proc. of SPIE*, 5113, p. 252.
- Hägele, D.; Oestreich, M.; Rühle, W.W., Nestle, N.; Eberl, K. (1998). Spin transport in GaAs. *Appl. Phys. Lett.*, 73, 1580.
- Hruška, M.; Kos, Š.; Crooker, S.A.; Saxena, A.; Smith, D.L. (2006). Effects of strain, electric, and magnetic fields on lateral electron-spin transport in semiconductor epilayers. *Phys. Rev. B*, 73, 075306.
- Ivchenko, E.L.; Lyanda-Geller, Y.B. & Pikus, G.E. (1990). Electric current of optically polarized and thermalized carriers. *Sov. Phys.-JEPT*, 71, 550.
- Jiang J.H. & Wu M.W. (2009). Electron spin relaxation in bulk III-V semiconductors from a fully microscopic kinetic spin Bloch equation approach. *Phys. Rev. B*, 79, 125206.
- Kikkawa, J.M. & Awschalom, D.D. (1998). Resonant Spin Amplification in n-Type GaAs. *Phys. Rev. Lett.*, 80, 4313.
- Kiselev, A.A. & Kim, K.W. (2000). Progressive suppression of spin relaxation in two-dimensional channels of finite width. *Phys. Rev. B*, 61, 13115.
- Litvinenko, K. L.; Leontiadou, M.A.; Li, J.; Clowes, S.K.; Emeny, M. T.; Ashley, T.; Pidgeon, C.R.; Cohen, L.F. & Murdin, B.N. (2010). Strong dependence of spin dynamics on the orientation of an external magnetic field for InSb and InAs. *Appl. Phys. Lett.*, 96, 111107.
- Mairesse, Y.; Haessler, S.; Fabre, B.; Higuette, J.; Boutu, W.; Breger, P.; Constant, E.; Descamps, D.; Mével, E.; Petit, S. & Salières, P. (2008). Polarization-resolved pump probe spectroscopy with high harmonics. *New J. Phys.*, 10, 025028.
- Mikhailov S.A. (2008). Electromagnetic response of electrons in graphene: Non-linear effects. *Physica E: Low-dimensional Systems and Nanostructures*, 40, 2626.
- Mikhailov S.A. (2009). Non-linear graphene optics for terahertz applications. *Microelectronics Journal*, 40, 712.
- Mishchenko, E.G. & Halperin, B.I. (2003). Transport equations for a two-dimensional electron gas with spin-orbit interaction. *Phys. Rev. B*, 68, 045317.
- Moreau, Ph.; Siegrist, M.R.; Brazis, R. & Raguotis R. (1999). Enhancement of the Third Harmonic Generation Efficiency in n-Type Si and InP by Cooling from Room Temperature to 80 K. *Mat. Science Forum*, 297-298, 315.
- Paget, D.; Lampel, G.; Sapoval, B. & Safarov V.I. (1977). Low field electron-nuclear spin coupling in gallium arsenide under optical pumping conditions. *Phys. Rev. B*, 15, 5780.

- Perry, M. & Krane, J.(1993). High-order harmonic emission from mixed fields. *Phys. Rev. A*,48, R4051.
- Nougier, J.P. (1994). Fluctuations and noise of hot carriers in semiconductor materials and devices. *IEEE Trans. Electr. Dev.*, 41, 2034.
- Persano Adorno, D.; Zarcone, M.& Ferrante, G. (2000). Far-Infrared Harmonic Generation in Semiconductors. A Monte Carlo Simulation. *Laser Physics*, 10, 310.
- Persano Adorno, D.; Zarcone, M.& Ferrante, G. (2001). Monte Carlo Simulation of Harmonic Generation in InP. *Laser Part. Beams*,19, 81.
- Persano Adorno, D.; Zarcone, M.& Ferrante, G. (2003a). High-Order Harmonic Emission from Mixed Fields in n-type low-doped Silicon. *Laser Physics*,13, 270.
- Persano Adorno, D.; Zarcone, M.& Ferrante, G. (2003b). High harmonic generation by two color field-mixing in n-type low-doped GaAs. *Phys. Status Solidi C* ,0, 1488.
- Persano Adorno, D.; Zarcone, M.; Ferrante, G.; Shiktorov, P.; Starikov, E.; Gruzinskis, V.; Perez, S.; Gonzalez, T.; Reggiani, L.; Varani, L. & Vaissiere, J.C. (2004).Monte Carlo Simulation of high-order harmonics generation in bulk semiconductors and submicron structures. *Phys. Stat. Sol. C* ,1, 1367.
- Persano Adorno, D.; Capizzo, M.C. & Zarcone, M. (2007a).Monte Carlo Simulation of Harmonic Generation in GaAs structures operating under large-signal Conditions. *J. Comput. Electron.*,6, 27.
- Persano Adorno, D.; Zarcone, M.& Ferrante, G. (2007b). Generation of even harmonics of sub-THz radiation in bulk GaAs in the presence of a static electric field. *J. Comput. Electron.*,6, 31.
- Persano Adorno, D.; Capizzo, M.C. & Zarcone, M. (2008a). Changes of electronic noise induced by oscillating fields in bulk GaAs semiconductors. *Fluct. Noise Lett.*, 8, L11.
- Persano Adorno, D.; Pizzolato, N. & Spagnolo, B. (2008b). External noise effects on the electron velocity fluctuations in semiconductors. *Acta Phys. Pol. A*, 113, 985.
- Persano Adorno, D.; Pizzolato, N.& Spagnolo, B. (2009a). Noise influence on electron dynamics in semiconductors driven by a periodic electric field. *Journal of Statistical Mechanics: Theory and Experiment* , P01039-10.
- Persano Adorno, D.; Pizzolato, N.& Spagnolo, B. (2009b). Monte Carlo Study of Diffusion Noise Reduction in GaAs Operating under Periodic Conditions. *CP1129, Noise and Fluctuations, Proc. of the 20th International Conference (ICNF 2009)*, edited by M. Macucci and G. Basso,(American Institute of Physics) p 121.
- Persano Adorno, D.(2010). Polarization of the Radiation Emitted in GaAs Semiconductors Driven by Far-Infrared Fields *Laser Physics*, 20, 1061.
- Pershin, Y. (2005). Long-lived spin coherence states in semiconductor heterostructures. *Phys. Rev. B*,71, 155317.
- Pikus, G.E & Titkov, A.N. (1989) *Optical Orientation*, edited by Meyer F, Nauka (Leningrad).
- Pramanik, S.; Bandyopadhyay, S. & Cahay, M. (2003).Spin dephasing in quantum wires. *Phys. Rev. B*, 68, 075313.
- Romanov, Yu A.; Romanova, J. Yu; Mourokh, L.G. & Horing, N.J.M. (2004). Nonlinear properties of semiconductor superlattices in a biharmonic field. *Semicond. Sci. Technol.*,19, S80.
- Saikin, S.; Shen, M.; Cheng, M.C. & Privman, V. (2003). Semiclassical Monte Carlo model for in-plane transport of spin-polarized electrons in III V heterostructures. *J. Appl. Phys.*,94, 1769.

- Saikin, S. (2004). A drift-diffusion model for spin-polarized transport in a two-dimensional non-degenerate electron gas controlled by spin orbit interaction. *J. Phys.: Condens. Matter*, 16, 5071.
- Saikin, S.; Shen, M. & Cheng, M.C. (2006). Spin dynamics in a compound semiconductor spintronic structure with a Schottky barrier. *J. Phys.: Condens. Matter*, 18, 1535.
- Sanada,H.; Arata, I.; Ohno, Y.; Chen, Z.; Kayanuma,K.; Oka, Y.; Matsukura, F. & Ohno H.(2002). Relaxation of photoinjected spins during drift transport in GaAs. *Appl. Phys. Lett.*, 81, 2788.
- Shen, M.; Saikin, S.; Cheng, M.C. & Privman, V. (2004). Monte Carlo Modeling of Spin FETs Controlled by Spin-Orbit Interaction. *Mathematics and Computers in Simulation*, 65, 351.
- Shiktorov, P.; Starikov, E.; Gruzinskis, V.; Zarccone, M.; Persano Adorno, D.; Ferrante, G., (b), Reggiani, L.; Varani, L. & Vaissière, J.C. (2002a). Monte Carlo Analysis of the Efficiency of Tera-Hertz Harmonic Generation in Semiconductor Nitrides. *Phys. stat. sol. (a)* ,190, 271.
- Shiktorov, P.; Starikov, E.; Gruzinskis, V.; Reggiani, L.; Varani, L. & Vaissière, J.C. (2002b). Monte Carlo calculation of electronic noise under high-order harmonic generation. *Appl. Phys. Lett.*, 80, 4759.
- Shiktorov, P.; Starikov, E.; Gruzinskis, V.; Perez, S.; Gonzalez, T.; Reggiani, L.; Varani, L. & Vaissiere, J.C.(2003a). Monte Carlo simulation of threshold bandwidth for high-order harmonic extraction. *IEEE Trans. Electr. Dev.*, 50, 1171.
- Shiktorov, P.; Starikov, E.; Gruzinskis, V.; Pérez, S.; González, T.; Reggiani, L.; Varani, L. & Vaissière, J.C. (2003b). Upconversion of partition noise in semiconductors operating under periodic large-signal conditions. *Phys. Rev. B* , 67 165201.
- Song,X.; Gong, S.; Jin, S. & Xu Z.(2003). Two-color interference effects for ultrashort laser pulses propagating in a two-level medium. *Physics Letters A*, 319, 150.
- Spezia, S.; Persano Adorno, D.; Pizzolato, N.& Spagnolo, B. (2010). Temperature Dependence of Spin Depolarization of Drifting Electrons in n-type GaAs Bulks. *Acta Phys. Pol. B*, 41, 1171.
- Spezia, S.; Persano Adorno, D.; Pizzolato, N.& Spagnolo, B. (in press). Doping dependence of spin dynamics of drifting electrons in GaAs bulks. *Acta Phys. Pol. A*.
- Urban, M.; Nieswand, Ch.; Siegrist, M.R. & Keilmann,F. (1995). Intensity dependence of the third-harmonic-generation efficiency for high-power far-infrared radiation in n-silicon. *J. Appl. Phys.*,77, 981.
- Urban, M.; Siegrist, M.R.; Asadauskas, L.; Raguotis, R. & Brazis,R. (1996). A precise new method to evaluate Monte Carlo simulations of electron transport in semiconductors. *Appl. Phys. Lett.*,69, 1776.
- Varani, L.; Palermo, C.; De Vasconcelos, C.; Millithaler, J.F.; Vaissière, J.C.; Nougier, J.P.; Starikov, E.; Shiktorov, P. & Gruzinskis, V. (2005). Is It Possible To Suppress Noise By Noise In Semiconductors? *Proc.Int. Conf. on Unsolved Problems of Noise and Fluctuations*(American Institute of Physics) p 474.
- Vilar, J.M.G. & Rubi, J.M. (2001). Noise Suppression by Noise. *Phys. Rev. Lett.*, 86, 950.
- Walton, D.B. & Visscher, K. (2004). Noise suppression and spectral decomposition for state-dependent noise in the presence of a stationary fluctuating input. *Phys. Rev. E*, 69 051110-1.
- Wang, B.; Li, X. & Fu,P.(1999). Polarization effects in high-harmonic generation in the presence of static-electric field. *Phys. Rev. A*, 59, 2894.

- Weng, M.Q. & Wu, M.W. (2003). Kinetic theory of spin transport in n-type semiconductor quantum wells. *J. Appl. Phys.*, 93, 410.
- Weng, M.Q.; Wu, M.W. & Jiang, L. (2004). Hot-electron effect in spin dephasing in n-type GaAs quantum wells. *Phys. Rev. B*, 69, 245320.
- Wu, M.W. & Ning, C.Z. (2000). Dyakonov-Perel Effect on Spin Dephasing in n-Type GaAs. *Phys. Status Solidi B*, 222, 523.
- Wu, M.W.; Jiang, J.H. & Weng, M.Q. (2010). Spin dynamics in semiconductors. *Physics Reports*, 493, 61.
- Xia, K.; Niua, Y.; Li C. & Gong, S. (2007). Absolute phase control of spectra effects in a two-level medium driven by two-color ultrashort laser pulses. *Physics Letters A*, 361, 173.
- Yafet, Y. (1963) *Solid State Physics*, edited by F. Seitz and D. Turnbull, (New York) Academic, Vol. 14, p. 2.
- Zhang, P.; Zhou, J. & Wu, M.W. (2008). Multivalley spin relaxation in the presence of high in-plane electric fields in n-type GaAs quantum wells. *Phys. Rev. B*, 77, 235323.

# A Pearson Effective Potential for Monte-Carlo Simulation of Quantum Confinement Effects in nMOSFETs

Marie-Anne Jaud<sup>1</sup>, Sylvain Barraud<sup>1</sup>, Philippe Dollfus<sup>2</sup>,  
Jérôme Saint-Martin<sup>2</sup>, Arnaud Bournel<sup>2</sup> and Hervé Jaouen<sup>3</sup>

<sup>1</sup>CEA-LETI, MINATEC, 17 rue des Martyrs, 38054 Grenoble,

<sup>2</sup>Institut d'Electronique Fondamentale, CNRS UMR 8622,  
Bât. 220, Univ. Paris-Sud, 91405 Orsay,

<sup>3</sup>STMicroelectronics, 850 rue Jean Monnet, 38926 Crolles,  
France

## 1. Introduction

As MOSFETs are downscaled to nanometric dimensions, ultra-thin body devices are required for an optimal electrostatic channel control. In such devices, quantization effects are likely to have a large impact on both electrostatics and carrier transport properties. Consequently, to accurately investigate electron transport in ultimate MOSFET architectures, the usual semi-classical transport models can no longer be applied and new simulation tools accounting for quantum effects in the electron transport description are becoming of great relevance.

In the last few years, some works investigated the possibility to develop quantum models based on a particle description of transport. Given the strong analogy between Wigner and Boltzmann formalisms, the Monte-Carlo method commonly used for semi-classical transport simulation can be extended to the quantum case by considering the Wigner function as an ensemble of pseudo-particles (Shifren et al., 2003 ; Nedjalkov et al., 2004 ; Querlioz et al., 2006). This approach describes well the wave-like nature of particles and has been first applied to the one-dimensional (1D) simulation of double-barrier resonant structures. To treat quantization effects in an inversion channel, one may couple self-consistently the 1D Schrödinger equation solved along the confinement direction with the multi-subband Boltzmann transport in the source-to-drain direction including 2D scattering rates (Lucci et al., 2005 ; Saint-Martin et al., 2006). This mode-space approach properly accounts for quantization effects in ultra-thin double-gate devices but is computationally intensive and may be difficult to extend to other architectures. Recently, some works combining the two previous methods for studying quantum transport in ultra-scaled double-gate MOSFETs have been published (Sverdlov et al., 2005; Querlioz et al., 2007). Alternatives to the mode-space approach are the quantum corrected potential methods (Ferry et al., 2000; Akis et al., 2001; Li et al., 2002; Tang et al., 2003; Tsuchiya et al., 2003; Fan et al., 2004; Ahmed et al., 2005; Riolino et al., 2006; Jaud et al., 2006) which have been demonstrated as an efficient way for including quantization effects in a semi-classical

particle Monte-Carlo simulator. Among these techniques, the Gaussian Effective Potential (GEP) formulation (Ferry et al., 2000; Akis et al., 2001; Li et al., 2002; Palestri et al., 2005; Jaud et al., 2006) is of great interest because it is weakly sensitive to the particle noise inherent in Monte-Carlo simulation and it is an alternative to the Schrödinger-Poisson based effective potential (Fan et al., 2004) that requires to solving the Schrödinger's equation. As already reported in (Li et al., 2002; Palestri et al., 2005; Jaud et al., 2006), the GEP correction can accurately reproduce Schrödinger-Poisson (SP) integral quantities such as the total inversion charge but fails to correctly model the electron density profiles. The discrepancy between GEP and SP density profiles is particularly important close to the SiO<sub>2</sub>/Si interfaces. It is thus especially critical in ultra-thin double-gate structures where electron wave functions are affected by two such interfaces.

In this chapter, we demonstrate the ability of an original Effective Potential formalism to properly introduce the quantum confinement effects in a Monte-Carlo simulator, i.e. not only the electrostatics in long nMOS capacitors but also the electron transport in nanoscale nMOSFET devices. In section 2, we briefly outline the quantum corrected potential approach for Monte-Carlo simulation. Section 3 highlights and investigates the limitations of the usual Gaussian Effective Potential (GEP). This leads us to develop a novel Pearson Effective Potential (PEP) correction, whose detailed description, electrostatic validation on various MOS architectures and extension to source and drain areas are described in sections 4, 5 and 6, respectively. Section 7 compares the results obtained from semi-classical, GEP corrected, PEP corrected and multi-subband Monte-Carlo methods for an ultra-short double-gate nMOSFET at low and high drain voltage. Finally, the influence of quantum confinement effects on the drive current as a function of both the channel length and the silicon film thickness is discussed in section 8.

## 2. Quantum corrected potential approach

The quantum corrected potential concept has been first introduced by Madelung and Bohm (Madelung, 1926; Bohm, 1952). Its aim is to reproduce physical effects due to quantization by modifying the electrostatic potential responsible for the carrier movement. The flowchart of the quantum corrected Monte-Carlo algorithm together with an illustration of the potential and of the electron density as a function of the distance from an oxide/silicon interface along the confinement direction (referred to as  $x$ -axis) are presented in Fig. 1.

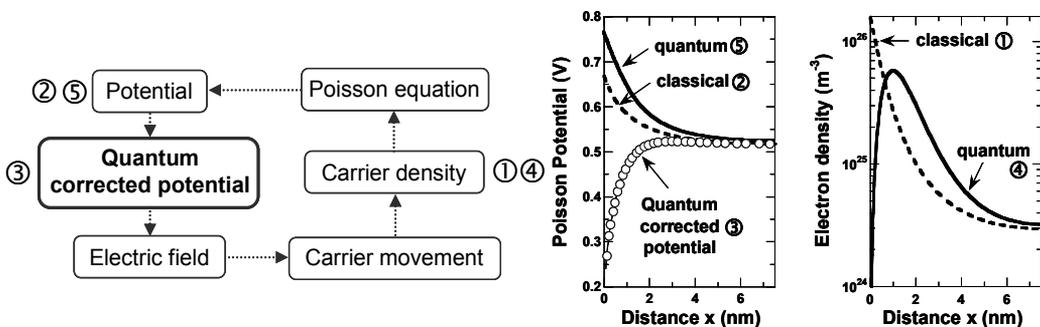


Fig. 1. Principle of the quantum corrected Monte Carlo simulation.

At first, the potential obtained from Poisson's equation solution is used to calculate the quantum corrected potential to be introduced in the Monte-Carlo algorithm for the calculation of carrier trajectories. The resulting quantum corrected potential generates an electric field that tends to repel carriers from the oxide/silicon interfaces in accordance with quantization effects. The carrier repulsion at interface is thus naturally included in the standard Monte-Carlo algorithm. As expected, the Poisson equation solution leads to a "quantum" potential which has a higher curvature than the "classical" potential. Finally, the self-consistency between quantum corrected potential and carrier movement is obtained from an iterative procedure. Within this approach, only the free-flight carrier trajectories are modified by the quantum correction. Scattering mechanisms are assumed to be identical to those of a conventional semi-classical Monte-Carlo approach.

### 3. Gaussian effective potential model

#### 3.1 Theoretical model

The effective potential formalism has been originally developed by Feynman (Feynman & Hibbs, 1965). It accounts for carrier non-locality by considering the finite size of the carrier wave-packet. As a result, a carrier is not only influenced by the local potential at its position but also by the neighboring potential distribution. The usual Gaussian Effective Potential (GEP) is defined along the confinement direction by the convolution of the Poisson potential with a Gaussian function representing the electron wave-packet (Feynman & Hibbs, 1965; Ferry et al., 2000):

$$\text{GEP}(x) = \frac{1}{\sqrt{2\pi} \sigma_x} \int_{-T_{\text{ox}}}^{T_{\text{Si}}+T_{\text{ox}}} V_P(x') \times \exp\left(-\frac{|x-x'|^2}{2\sigma_x^2}\right) dx' \quad (1)$$

where  $\sigma_x$  is the standard deviation of the Gaussian function,  $T_{\text{Si}}$  the silicon film thickness,  $T_{\text{ox}}$  the oxide thickness and  $V_P(x')$  the Poisson potential. As explained in (Jaud et al., 2006), the GEP is calculated using a Fourier transform method. Accordingly, to apply appropriate boundary conditions to the Poisson potential on the oxide areas and to avoid data corruption by convolution in equation (1), "Padding regions" (by reference to signal processing techniques) are used on the edge of the device. The parameter  $E_B = 3.1$  eV is defined at the  $\text{SiO}_2/\text{Si}$  interfaces to represent the oxide barrier height for electrons and satisfies  $V_{\text{oxide}} = V_P - E_B$ .

#### 3.2 Results and discussion

As described in (Jaud et al., 2006), we have implemented the GEP correction in the framework of a Monte-Carlo code (MONACO) (Saint-Martin et al., 2004) that uses an analytical conduction-band structure of silicon considering six ellipsoidal nonparabolic  $\Delta$  valleys. Double-gate nMOS capacitors with a channel doping  $N_A = 10^{16} \text{ cm}^{-3}$  and an oxide thickness  $T_{\text{ox}} = 1$  nm have been simulated. Self-consistent Monte-Carlo simulations corrected by the GEP have been performed for a large range of silicon thicknesses ( $5 \text{ nm} \leq T_{\text{Si}} \leq 20 \text{ nm}$ ) together with a perpendicular effective field  $E_{\text{eff}}$  varying from  $10^5 \text{ V.cm}^{-1}$  to  $10^6 \text{ V.cm}^{-1}$ . In accordance with (Akis et al., 2001; Palestri et al., 2005), the standard deviation of the Gaussian function is chosen to be equal to  $\sigma_x = 0.5$  nm. Considering the results from SP simulations including the 2-fold and 4-fold valleys with 10 energy levels for each valley as reference, Fig. 2a shows the error on the inversion charge induced by the GEP correction.

Fig. 2b compares the electron density resulting from the GEP correction with the one resulting from SP simulation for  $T_{Si} = 10$  nm. The GEP formalism is well-known and has been proved to be useful to describe “electrostatic quantum effects” (Ferry et al., 2000 ; Akis et al., 2001 ; Li et al., 2002 ; Palestri et al., 2005). However, errors higher than 10% on the inversion charge are observed at  $E_{eff} = 10^5$  V.cm<sup>-1</sup>. At this low effective field, a decrease of the silicon thickness yields a noticeable increase of the inversion charge error (cf. Fig. 2a). Moreover, in agreement with (Li et al., 2002 ; Palestri et al., 2005), one can observe in Fig. 2b that the results obtained from Monte-Carlo simulation corrected by the GEP show an overestimated carrier repulsion at the SiO<sub>2</sub>/Si interfaces. This is due to the fact that the electron wave-packet is systematically represented by a unique Gaussian function, defined by a standard deviation  $\sigma_x$  and an average position  $R_p$ , all along the silicon film thickness. Close to the SiO<sub>2</sub>/Si interfaces, this description is not realistic with regard to SP results. The inability of the Gaussian function to represent the electron wave-packet has been clearly highlighted in (Jaud et al., 2006) using a methodology based on a design-of-experiments. It has been proved impossible to find out any values of  $E_B$  and  $\sigma_x$  likely to properly reproduce the SP carrier density profile.

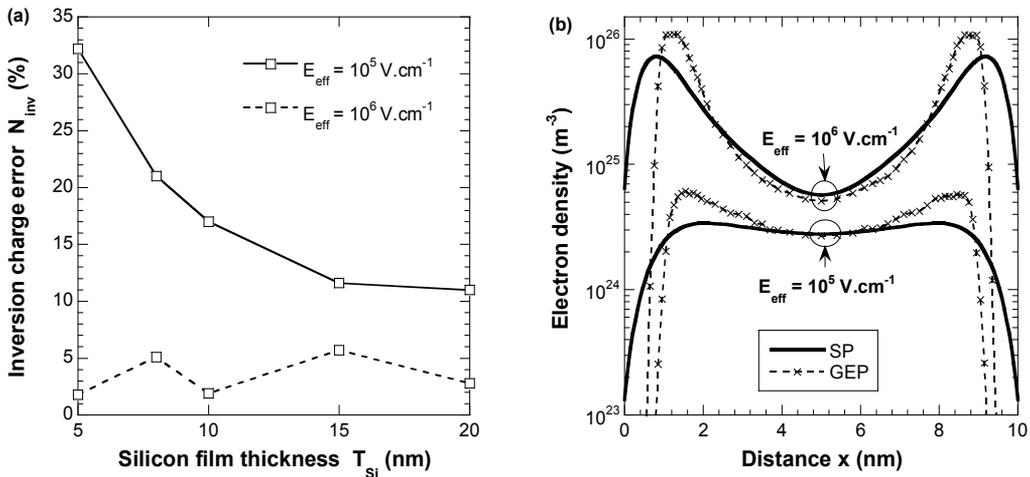


Fig. 2. (a) Inversion charge error in GEP correction (with standard parameters  $E_B = 3.1$  eV and  $\sigma_x = 0.5$  nm) as a function of the silicon film thickness of double-gate nMOS capacitors for  $10^5$  V.cm<sup>-1</sup> (solid line) and  $10^6$  V.cm<sup>-1</sup> (dotted line) perpendicular effective fields. (b) Electron density as a function of the distance in the confinement direction in a double-gate nMOS capacitor with  $T_{Si} = 10$  nm using Schrödinger-Poisson (SP - solid lines) and Monte-Carlo corrected by the GEP (GEP - cross dotted lines) models.

## 4. Pearson effective potential model

### 4.1 General principle

The previous study based on the GEP correction leads us to propose a new Effective Potential formalism where the electron wave-packet description is improved. The Gaussian function is replaced by a more realistic function based on the shape of the squared modulus of the first level Schrödinger's wave function  $|\psi_0|^2$  and carefully calibrated so as to reproduce the electron density profiles resulting from SP simulations considering 10 energy

levels. Before calibrating our new function, we first have (i) to choose a well-suited function to reproduce the different possible shapes of  $|\psi_0|^2$ ; (ii) to identify the parameters responsible for the main characteristics of the shape of  $|\psi_0|^2$ , i.e., to determine the dependences to be given to the new electron wave-packet description. This will lead us to define our novel effective potential formulation.

### Electron wave-packet's description

To well describe the various shapes of  $|\psi_0|^2$ , the new function has to verify the two following conditions: (i) to be a generalization of the Gaussian distribution and (ii) to be possibly asymmetrical. The *Pearson type IV distribution*, often used for the description of doping implantation profiles, fully satisfies these conditions. It is defined by its first four moments which are related to the average position ( $R_p$ ), the standard deviation ( $\sigma_p$ ), the skewness ( $\gamma$ ) and the kurtosis ( $\beta$ ) of the distribution, respectively (Selberherr, 1984; Sze, 1988) (see 11. Appendix). Fig. 3 illustrates the influence of each Pearson IV parameter. The skewness and the kurtosis are a measure of the *asymmetry* and *peakedness* of the distribution function, respectively. A positive, respectively negative, value of the skewness results in a maximum of the distribution on the left, respectively on the right, of its average position (cf. Fig. 3b). We can note that a Gaussian function is a particular Pearson IV distribution defined by  $\gamma=0$  and  $\beta=3$ .

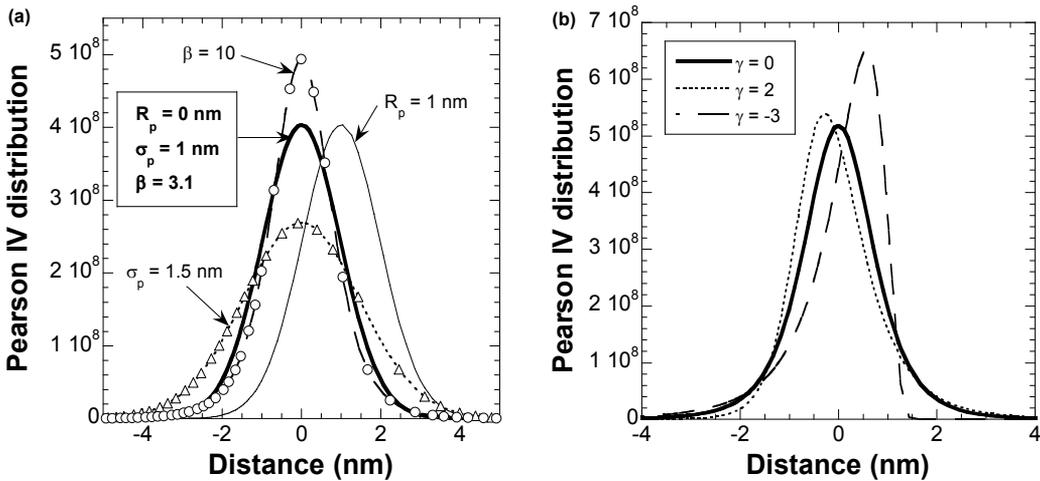


Fig. 3. Pearson IV distributions. (a)  $R_p = 0$  nm,  $\sigma_p = 1$  nm,  $\gamma = 0$ ,  $\beta = 3.1$  (solid heavy line) /  $R_p = 1$  nm,  $\sigma_p = 1$  nm,  $\gamma = 0$ ,  $\beta = 3.1$  (solid line) /  $R_p = 0$  nm,  $\sigma_p = 1.5$  nm,  $\gamma = 0$ ,  $\beta = 3.1$  (open triangles) /  $R_p = 0$  nm,  $\sigma_p = 1$  nm,  $\gamma = 0$ ,  $\beta = 10$  (open circles). (b)  $R_p = 0$  nm,  $\sigma_p = 1$  nm,  $\beta = 30$ .

### Electron wave-packet's dependences

It is well-known that the shape of  $|\psi_0|^2$  is primarily influenced (i) by the potential profile in the confinement direction and (ii) by the silicon film thickness. Therefore, so as to realistically describe the particle wave-packet, Pearson IV parameters should depend (i) on the local electric field  $E_x$  in the confinement direction, calculated as the derivative of the potential obtained from Poisson's equation in the confinement direction and (ii) on the silicon film thickness  $T_{Si}$ . This way, the influence of parameters such as  $T_{ox}$ ,  $N_A$  or gate voltage is implicitly taken into account through the  $E_x$ -dependence.

### Pearson Effective Potential formulation

As in the GEP approach, our PEP formulation is based on the convolution of the Poisson potential by a Pearson IV function representing the non zero-size of the electron wave-packet (Feynman & Hibbs, 1965; Ferry et al., 2000). For a double-gate structure it is defined (1D) as:

$$\text{PEP}(x) = \int_{-T_{\text{ox}}}^{T_{\text{Si}}+T_{\text{ox}}} [V_p(x') * \text{Pearson IV}(R_p(E_x, T_{\text{Si}}) - x')] dx' \quad (2)$$

where  $V_p(x')$  is the potential energy,  $T_{\text{Si}}$  and  $T_{\text{ox}}$  are the silicon film and oxide thicknesses, and  $E_x$  is the local electric field in the confinement direction.

### 4.2 Calibration

To calibrate the four moments of the Pearson IV distribution, the Schrödinger-Poisson equations considering 10 energy levels have been solved self-consistently for double-gate nMOS capacitors with silicon film thickness varying from 5 nm  $\leq T_{\text{Si}} \leq 20$  nm and for a large range of effective fields ( $10^5 \text{ V.cm}^{-1} \leq E_{\text{eff}} \leq 10^6 \text{ V.cm}^{-1}$ ). Indeed, double-gate capacitors with  $T_{\text{Si}}$  less than 5 nm are not very realistic for actual technological purposes and the chosen range of effective fields is similar to the values used for the effective mobility extraction in the inversion layer of long-channel devices.

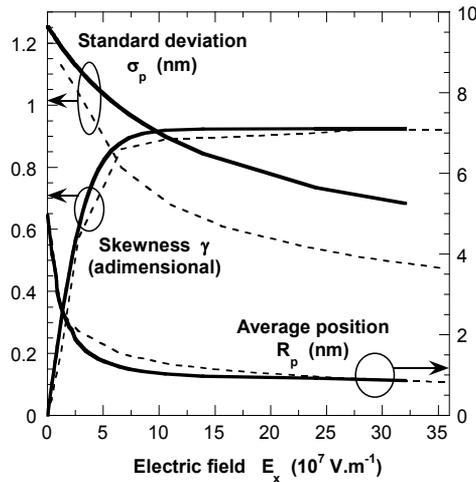


Fig. 4.  $R_p$ ,  $\sigma_p$  and  $\gamma$  as a function of the electric field  $E_x$  in the confinement direction extracted from the squared modulus of the first level Schrödinger's wave function (dotted lines) and defining the Pearson IV distribution of the PEP model (solid lines) for  $T_{\text{Si}} = 10$  nm.

For each device and effective field, the interfacial electric field, the squared modulus of the first level Schrödinger's wave function  $|\psi_0|^2$  and the electron density profile have been extracted. Then, each of the first four theoretical moments of  $|\psi_0|^2$  has been calculated as a function of the interfacial electric field and of the silicon film thickness. Thereafter, the terminology "theoretical values" refers to these moment values deduced from SP  $|\psi_0|^2$  functions. In the case of a 10 nm film thickness double-gate capacitor, the theoretical values

of the average position with respect to the oxide-silicon interface, the standard deviation and the skewness are plotted in dotted lines as a function of the interfacial electric field on Fig. 4. When decreasing the electric field, the average position is farther away from oxide/silicon interface, the standard deviation is greater and the skewness is smaller, which is in accordance with less pronounced quantum confinement effects. The first four moments defining the Pearson IV distributions were calibrated using appropriate functions both to fit theoretical values of  $|\psi_0|^2$  as closely as possible and to reproduce SP electron density profiles. The solid lines of Fig. 4 shows the calibration results of average position, standard deviation and skewness obtained for a double-gate capacitor of 10 nm film thickness. Moreover, for this structure in inversion regime, some Pearson IV distributions associated with various carrier positions in the silicon film as well as the first four moments of the Pearson IV are plotted on Fig. 5 along the confinement direction.

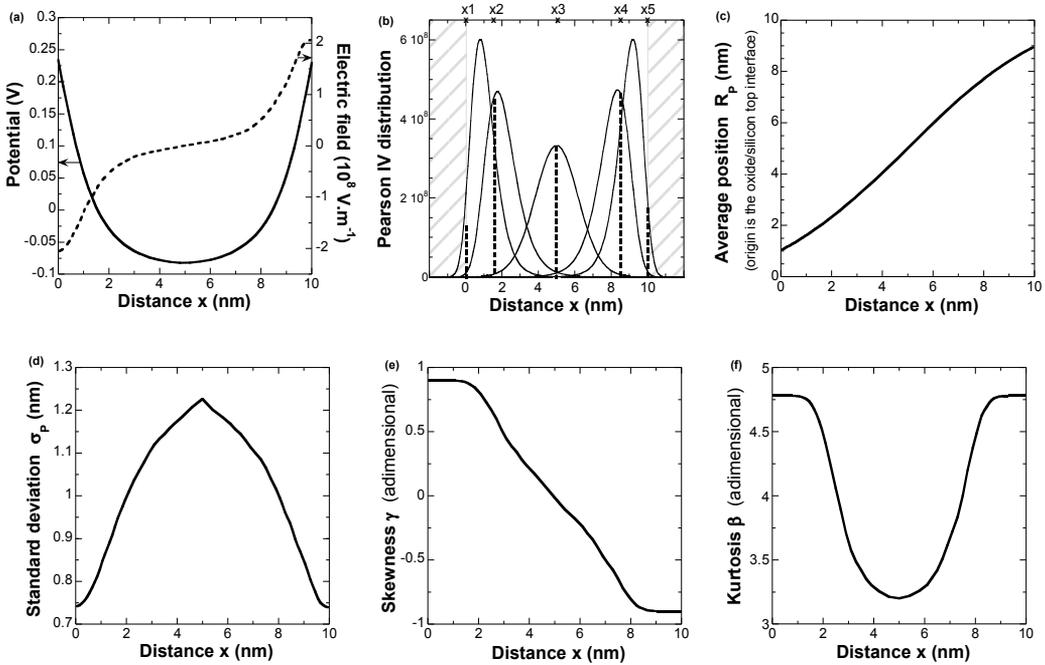


Fig. 5. Poisson potential and local electric field (a), Pearson IV distributions representing the electron wave-packets associated to various electron positions (symbolized by dotted lines) (b),  $R_p$  (c),  $\sigma_p$  (d),  $\gamma$  (e) and  $\beta$  (f) as a function of the distance along the confinement direction for a  $T_{Si} = 10$  nm double-gate nMOS capacitor in inversion regime.

Now we describe in more details the fitting procedure. The expressions of Pearson IV moments as a function of  $E_x$  and  $T_{Si}$  together with the resulting fitting parameters are given in (11. Appendix).

- For the definition of the average position ( $R_p$ ), the position of the oxide/silicon top interface is taken as reference. As a function of  $E_x$  and  $T_{Si}$ ,  $R_p$  is chosen to fit the theoretical values (cf. Fig. 4) while ensuring that (i) in the case of a zero electric field the average position  $R_p$  is equal to the particle position and (ii) the  $R_p$  evolution along the

confinement direction  $x$  is continuous and regular, a necessary condition for the numerical stability of the algorithm. We can note in Fig. 5c that the average position of the wave-packet of a particle located at the oxide/silicon interface is at about 1 nm apart from this interface, which prevents from unrealistic wave-packet penetration in the oxide layer.

- The standard deviation ( $\sigma_p$ ) has been considered as the unique adjustable parameter; i.e. it is not chosen to accurately fit the "theoretical value" but to reproduce the SP electron density profiles. It is explained by the fact that, from SP solution, a weak penetration of the wave-functions in the oxide layer leads to a strong carrier repulsion. In contrast, in Monte-Carlo simulation corrected by an effective potential, a weak penetration of the distribution function assimilated to the particle wave-packet in the oxide layer originates a weak repulsive electric field close to the oxide/silicon interfaces, which therefore results in a weak carrier repulsion. That is why the standard deviation of the Pearson IV is not taken identical to the theoretical one but is generally taken slightly higher (cf. Fig. 4). More precisely,  $\sigma_p$  is chosen so that the Pearson penetration into the oxide layer induces a repulsive electric field which correctly reproduces electron density profile from SP simulation including several subbands.
- The skewness ( $\gamma$ ) of the Pearson IV distribution has been chosen by fitting the theoretical one (cf. Fig. 4). The sign of the electric field determines the sign of the skewness (cf. Fig. 5e).
- The kurtosis ( $\beta$ ) is arbitrarily calculated as a function of the skewness  $\gamma$  so as to be minimal and as close as possible to the Gaussian value (Selberherr, 1984; Sze, 1988).

Finally, this calibration procedure has allowed us to determine equations defining  $R_p$ ,  $\sigma_p$  and  $\gamma$  as a function of  $E_x$  and  $T_{Si}$  as well as  $\beta$  as a function of  $\gamma$  (see 11. Appendix). This way, for each carrier position in the confinement direction, the associated Pearson IV distribution is fully defined (cf. Fig. 5b). It can be noted that the Pearson IV representing the wave-packet of a particle located at  $SiO_2/Si$  interfaces ( $x=0=x_1$  and  $x=T_{Si}=x_5$ ) is centred on  $R_p \neq x$  and presents a noticeable asymmetry  $\gamma \neq 0$ . On the other hand, for a particle located at  $x=T_{Si}/2=x_3$ , the Pearson IV looks like a Gaussian function ( $\gamma=0$ ) and is centred on  $R_p=x=T_{Si}/2$ . With our new approach, all along the silicon film thickness and particularly close to the  $SiO_2/Si$  interfaces, the particle wave-packet representation is clearly more realistic than a Gaussian distribution. Moreover, since we have calibrated our PEP correction so as to reproduce electron density profiles resulting from SP calculation including 10 energy levels, one can say that our PEP correction integrates the description of valleys and of their associated subbands. However, this technique cannot include the confinement-induced redistribution of electrons among the different valleys as can be done in the Schrödinger-based correction method (Fan et al., 2004).

#### 4.3 PEP calculation flowchart

The generic flowchart of the PEP calculation is presented in Fig. 6. As for the GEP correction, (i) the PEP correction has been implemented in the framework of a Monte-Carlo code (MONACO) (Saint-Martin et al., 2004), (ii) the parameter  $E_B = 3.1$  eV is defined at the  $SiO_2/Si$  interfaces and satisfies  $V_{oxide} = V_P - E_B$ .  $E_x$  and  $T_{Si}$  being known, a set of four parameters ( $R_p$ ,  $\sigma_p$ ,  $\gamma$ ,  $\beta$ ) defining a Pearson IV distribution is calculated at each grid node of the structure as described in the previous section. Let us recall that the solution of Schrödinger's equation is not required for the PEP calculation. The Pearson IV determination

only needs the knowledge of calibrated parameters. The Pearson Effective Potential is then calculated at each position “x” as the integral (see eq. 2) of the product of the Poisson Potential with the associated Pearson IV distribution. Due to the different shapes of the Pearson IV distributions to be considered all along the silicon film thickness, the PEP correction can no longer be performed by a Fourier transform method as in the case of the GEP correction. It is now calculated using a Gaussian quadrature numerical integration method (Dhatt et al., 2005).

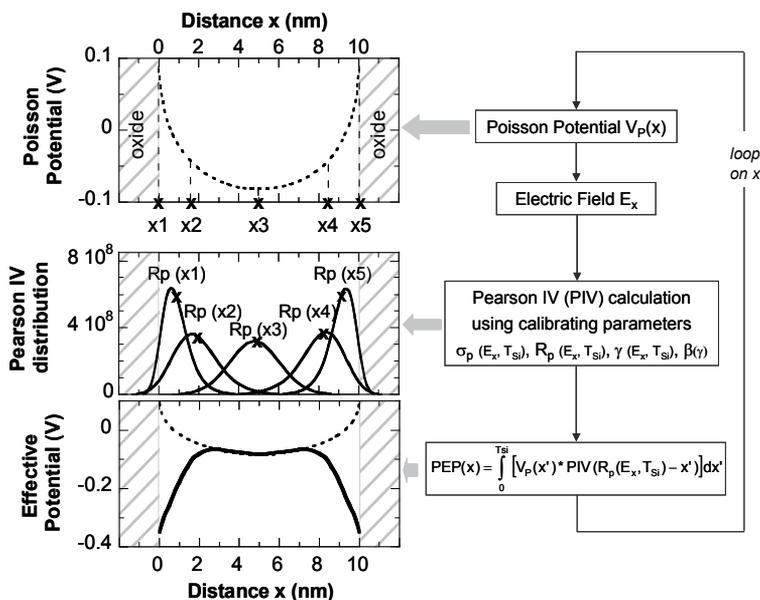


Fig. 6. Flowchart of the Pearson Effective Potential calculation illustrated by results on a double-gate device with  $T_{Si} = 10$  nm.

### 5. Pearson Effective Potential electrostatics validation

To validate the original PEP formulation, self-consistent simulations have been performed for several device architectures (double-gate, Silicon On Insulator (SOI) and bulk). Results of Monte-Carlo simulation corrected by the PEP model are compared with that obtained from SP calculation and GEP-corrected Monte Carlo simulation (with the value  $\sigma_x = 0.5$  nm, as in (Akis et al., 2001 ; Palestri et al., 2005)). Because of confinement effects close to both  $SiO_2/Si$  interfaces, the double-gate nMOS architecture is one of the most critical devices to be tested to assess and demonstrate the ability of our PEP correction to reproduce the SP simulation results. The electron density profiles extracted from double-gate nMOS capacitors with 10 nm silicon film thickness and for a large range of effective fields ( $10^5 V.cm^{-1} \leq E_{eff} \leq 10^6 V.cm^{-1}$ ) are shown in Figure 7a. While the electron density profiles calculated with the GEP correction are clearly unrealistic close to the  $Si/SiO_2$  interfaces due to an unsuitable description of the particle wave-packet, those obtained by the PEP correction agree very well with SP results. Fig. 7b compares the Poisson potential resulting from the PEP correction (open circles) with that resulting from SP simulation (solid line). An excellent agreement is obtained between both approaches. The Poisson potential resulting from semi-

classical Monte-Carlo simulation (dotted line) and the Pearson Effective Potential which is actually responsible for the carrier movement (open squares) are also plotted in Fig. 7b. As expected the “quantum” Poisson potential exhibits a higher curvature than the “classical” one. Same results have been shown for double-gate nMOS capacitors with an oxide thickness  $T_{ox}$  varying from 0.5 nm up to 2 nm and a silicon film thickness  $T_{Si}$  ranging from 20 nm down to 5 nm without any changes in the Pearson IV parameters (Jaud et al., 2007a).

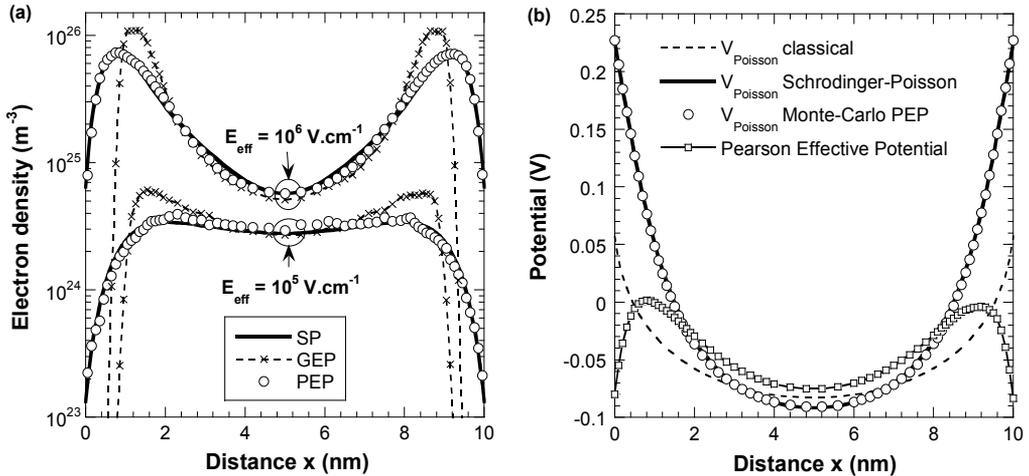


Fig. 7. (a) Electron density as a function of the distance in the confinement direction in a double-gate nMOS capacitor with  $T_{Si} = 10$  nm,  $T_{ox} = 1$  nm,  $N_A = 10^{16}$  cm $^{-3}$  and using SP (solid lines), GEP (cross dotted lines) and PEP (open circles) models. (b) Self-consistent Poisson Potential resulting from semi-classical (dotted line), SP (solid heavy line), Monte-Carlo with PEP correction (open circles) simulations and effective potential (PEP – open squares) as a function of the distance along the confinement direction extracted from the same capacitor.

Results obtained for a 5 nm silicon oxide thickness SOI capacitor and bulk nMOS capacitor with a channel doping  $N_A = 10^{18}$  cm $^{-3}$  and an oxide thickness  $T_{ox} = 1$  nm are presented in Fig. 8. The simulations have been performed using the same calibrated parameters as for the double-gate structure. The electron density resulting from the PEP correction still properly reproduces SP results. Finally, the ability of the GEP and PEP quantum corrections to conserve the total inversion charge  $N_{inv}$  for double-gate, SOI and bulk devices is gathered in Table 1. The results of SP simulations are taken as reference. At high effective field, the total inversion charge  $N_{inv}$  is accurately reproduced with both approaches. In contrast, at low effective field, the PEP correction generates an error of more than 10% lower than that induced by the GEP. Thus, besides reproducing accurately the SP electron density profiles, the PEP correction also leads to inversion charge errors at the worst equal to the GEP ones or even considerably reduced.

All these results highlight that the PEP correction is well-suited to predict electrostatic quantum confinement effects in ultimate bulk, SOI or double-gate nMOS devices with various  $T_{Si}$ ,  $T_{ox}$ ,  $N_A$  and gate bias without any additional calibration. This “universality” mainly results from a judicious calibration of Pearson IV parameters as a function of the local electric field in the confinement direction and of the silicon film thickness.

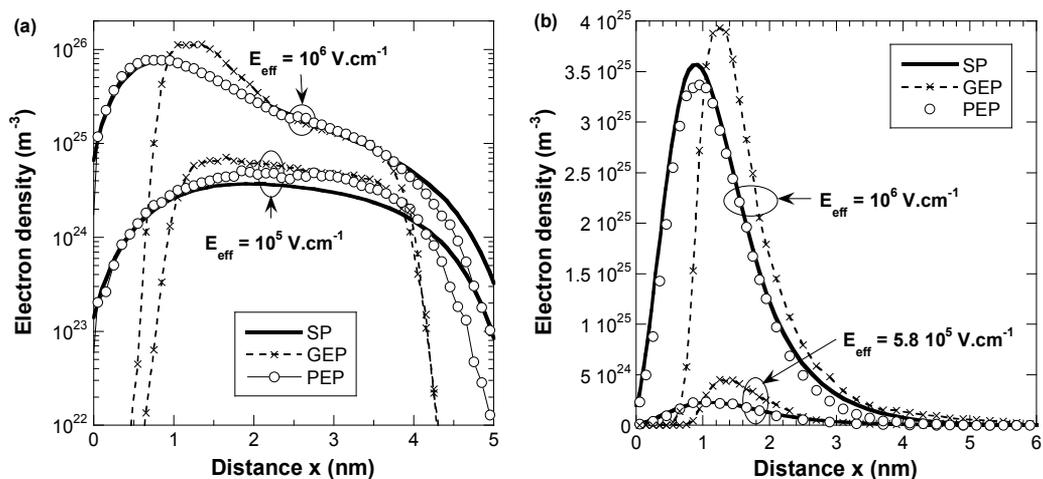


Fig. 8. Electron density as a function of the distance in the confinement direction in (a) a SOI nMOS capacitor with  $T_{Si} = 5 \text{ nm}$ ,  $T_{ox} = 1 \text{ nm}$ ,  $N_A = 10^{16} \text{ cm}^{-3}$  (b) a bulk nMOS capacitor with  $T_{ox} = 1 \text{ nm}$ ,  $N_A = 10^{18} \text{ cm}^{-3}$  and using SP (solid lines), GEP (cross dotted lines) and PEP (open circles) models.

nMOS capacitor			Monte-Carlo GEP		Monte-Carlo PEP	
Device	$T_{Si}$ (nm)	$T_{ox}$ (nm)	Low $E_{eff}$	High $E_{eff}$	Low $E_{eff}$	High $E_{eff}$
DG	20	1	11.0	2.8	0.7	0.4
DG	15	1	11.6	5.7	0.2	2.7
DG	10	1	17.1	1.9	4.8	1.6
DG	8	1	21.0	5.1	6.8	0.8
DG	5	1	32.2	1.8	21.6	3.9
DG	10	2	13.3	1.1	2.3	1.1
DG	10	0.5	23.2	4.1	5.6	2.7
SOI	10	1	24	3	1.6	0.7
SOI	5	1	35	2.3	23	2.3
Bulk		1	23.5	13.1	7.87	9.7

Table 1. Inversion charge error (in percentage) for various nMOS capacitors. Low  $E_{eff}$  corresponds to  $10^5 \text{ V.cm}^{-1}$  for double-gate (DG) and SOI devices and to  $5.8 \times 10^5 \text{ V.cm}^{-1}$  for bulk devices. High  $E_{eff}$  corresponds to  $10^6 \text{ V.cm}^{-1}$ .

## 6. Quantum correction for Monte-Carlo device simulation

### 6.1 Simulated device

The simulated device is a double-gate nMOSFET with a channel length  $L_C = 10 \text{ nm}$ , a  $\text{SiO}_2$  oxide thickness  $T_{ox} = 1.1 \text{ nm}$  and a silicon thickness  $T_{Si} = 5 \text{ nm}$ . Not only the channel but also the source and drain regions are covered with  $\text{SiO}_2$  oxide material. The source and drain regions are uniformly doped to  $10^{20} \text{ cm}^{-3}$  and the P-type residual doping level in the channel

is  $10^{15} \text{ cm}^{-3}$ . The metallic gate work function ( $\phi_M = 4.56 \text{ eV}$ ) corresponds to midgap material. The scattering mechanisms included in the model are the acoustic intravalley phonon scattering, three  $f$  and three  $g$  intervalley phonons scattering, and the electron-impurity scattering. In all simulations, the phonon scattering is computed via bulk-phonons using the same coupling constants that for 2D and 3D electron gas (Saint-Martin et al., 2006). To make easier the comparison between semi-classical, quantum corrected and multi-subband Monte-Carlo simulations with strictly similar scattering models, surface roughness scattering is not included here. Finally, degeneracy effects are not included in this work.

## 6.2 Treatment of quantization effects in source and drain

The simulation of the device presented above requires to considering quantum correction not only in the channel area but also in the source and drain areas where quantum confinement between oxide barriers also occurs. In these source and drain areas, and more generally in the case of a quasi-flat band potential profile in the confinement direction, the PEP correction consists in preserving the Gaussian distribution characteristic of the flat-band regime, but with an average position evolving at a small distance around the middle of the silicon film. This distance and the standard deviation of the Gaussian distribution are calibrated only as a function of the silicon film thickness.

To study the transition between the flat-band and the usual PEP corrections, the electron density profiles in the source-to-drain direction (referred to as  $y$ -axis) of an nMOS capacitor resulting from SP and PEP simulations are shown in Fig. 9. No discontinuity is observed at the source/channel junction. Moreover, a very good agreement is conserved between SP and PEP electron density profiles.

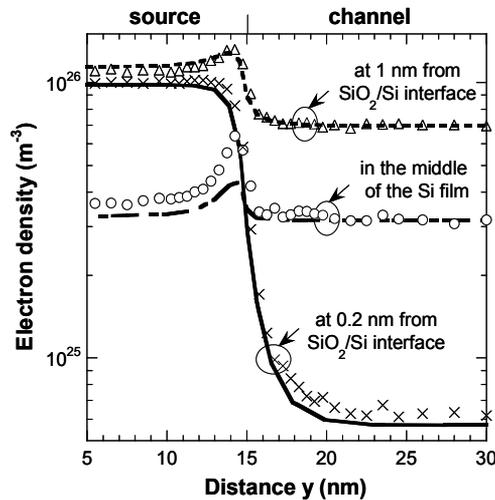


Fig. 9. Electron density profiles extracted in different slices along the transport direction close to the source/channel junction of a double-gate nMOS capacitor in inversion regime for SP (lines) and PEP model (symbols).  $T_{\text{Si}} = 10 \text{ nm}$ .

## 6.3 Boundary conditions at source and drain contacts

In a usual semi-classical Monte-Carlo approach, the charge neutrality conditions on plane ohmic contact are applied, that is not well-suited with quantum confinement effects in

source and drain areas. Thus, the following procedure has been adopted. The potential profile along  $x$  axis in the middle of the drain area (where boundary conditions do not affect the potential) is applied to the drain contact for Poisson's equation solution. The corresponding boundary condition, i.e. the potential profile applied at the source contact, is the same as at the drain contact but shifted by the drain to source voltage.

To conclude this section, thanks to the 1D PEP correction extended to the flat-band regime with the well-suited boundary conditions, the carrier quantum confinement is correctly described in the full nMOSFET structure.

## 7. PEP corrected vs. multi-subband Monte-Carlo

In this section the results obtained from semi-classical, GEP corrected, PEP corrected and multi-subband Monte-Carlo are compared. As for GEP and PEP, Multi-Subband Monte-Carlo (MSMC) approach (Saint-Martin et al., 2006) has been implemented in the framework of the MONACO code. Contrary to the GEP and PEP corrections, the MSMC solves the Schrödinger equation in the confinement direction and considers that all carriers are confined in a 2D gas. As a consequence, multi-subband (respectively GEP and PEP corrected) Monte-Carlo assumes 2D (respectively 3D) scattering rates and carrier movement. The MSMC approach is based on the mode-space approximation of decoupled 2D subbands only coupled by inter-subband scattering. This approximation is proved correct for ultra-thin double-gate structures ( $T_{Si} < 10$  nm) (Sverdlov et al., 2005). It may become questionable for structures where the subband coupling should be considered in the Schrödinger equation as in (Bulk or SOI) single-gate devices, and it cannot be easily applied in thicker devices since a dramatically large number of subbands and/or a tricky coupling with a 3D continuum of states should be required. In the present work, the MSMC approach has been performed for the 5 nm silicon film thickness double-gate nMOSFET previously described. In the following, electrical characteristics and then microscopic quantities are carefully compared at low and high drain voltages.

It should be noted that the GEP corrected, PEP corrected and multi-subband Monte-Carlo methods induce a computation-time multiplied by 2, 10 and 30 compared to that of semi-classical Monte-Carlo, respectively.

### 7.1 Current-voltage characteristics

The electrical output characteristics  $I_{DS}(V_{GS})$  calculated at  $V_{DS}=0.05$  V and  $I_{DS}(V_{DS})$  calculated at  $V_{GS}=1.2$  V resulting from semi-classical, GEP corrected, PEP corrected and multi-subband Monte-Carlo simulations are shown in Fig. 10. This figure demonstrates the limitations of the GEP correction to properly include quantization effects: it is not only unable to accurately reproduce electrostatic quantum confinement effects but it also over-corrects the current, which yields an underestimation of the drain current by more than 10% with respect to the MSMC results used as reference. Accordingly, the GEP correction is no longer considered in this work. In contrast, the PEP correction provides excellent results.

The total electron charge extracted at  $V_{DS}=0.05$  V and  $V_{DS}=0.7$  V as a function of the gate voltage and resulting from semi-classical, PEP corrected and multi-subband Monte-Carlo simulations is plotted in Fig. 11a. At the same drain voltages and at  $V_{GS}=1.2$  V, Fig. 11b represents the electron charge all along the device in the transport direction. It is remarkable that for all quantities plotted in Figs. 10-11 the PEP results fit in very well with multi-

subband ones. In Fig. 10, we observe a reduction of drive current when quantum confinement effects are included (by 6.2% with PEP at  $V_{GS}=1.2\text{ V}$  and  $V_{DS}=0.7\text{ V}$ ). It is mainly explained by a smaller effective gate capacitance due to carrier repulsion at the  $\text{SiO}_2/\text{Si}$  interfaces (cf. Fig. 11) inducing a reduction of inversion charge at given bias.

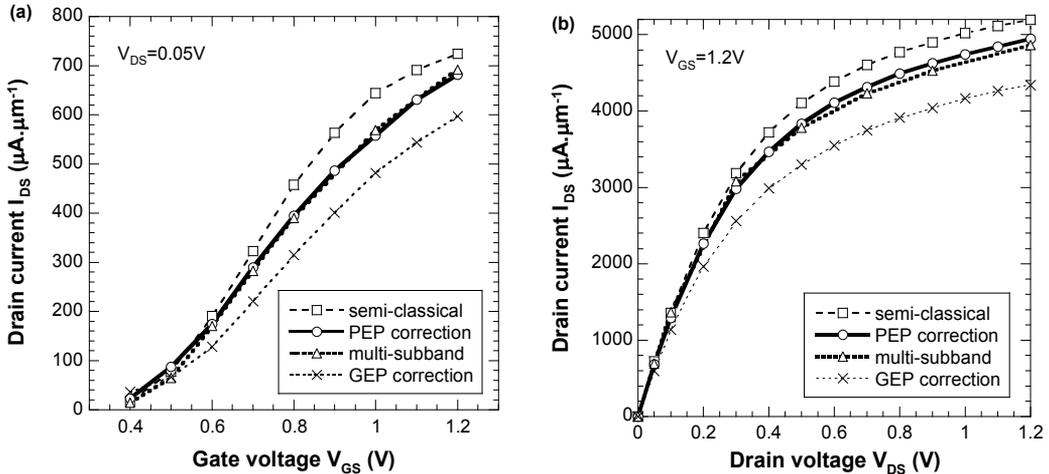


Fig. 10. Output  $I_{DS}$ - $V_{GS}$  characteristics at  $V_{DS} = 0.05\text{ V}$  (a) and output  $I_{DS}$ - $V_{DS}$  characteristics at  $V_{GS} = 1.2\text{ V}$  (b) resulting from semi-classical (squares), PEP corrected (circles), multi-subband (triangles) and GEP corrected (crosses) Monte-Carlo simulations.

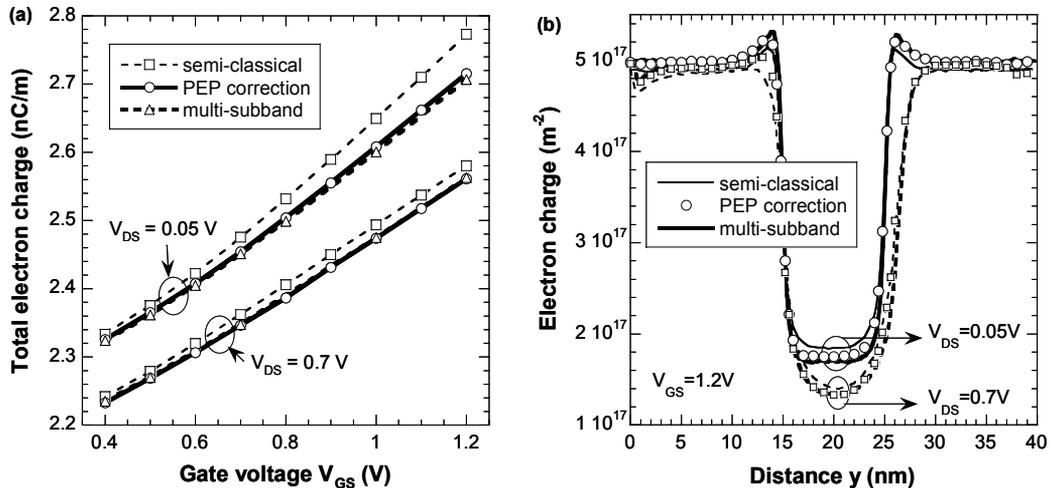


Fig. 11. (a) Total electron charge as a function of the gate voltage  $V_{GS}$  for  $V_{DS} = 0.05\text{ V}$  and  $V_{DS} = 0.7\text{ V}$  resulting from semi-classical (squares), PEP corrected (circles) and multi-subband (triangles) Monte-Carlo. (b) Electron charge (electron density integrated over the silicon film thickness) along the transport direction for  $V_{DS} = 0.05\text{ V}$  and  $V_{DS} = 0.7\text{ V}$  resulting from semi-classical (thin lines), PEP corrected (symbols) and multi-subband (thick lines) Monte-Carlo.  $V_{GS} = 1.2\text{ V}$ .

### 7.2 Microscopic quantities

Cartographies of electron density obtained by PEP corrected Monte-Carlo simulations at  $V_{DS} = 0.05$  V and  $V_{GS} = 1.2$  V and at  $V_{DS} = 0.7$  V and  $V_{GS} = 0.8$  V are shown in Figs. 12a and 13a, respectively.

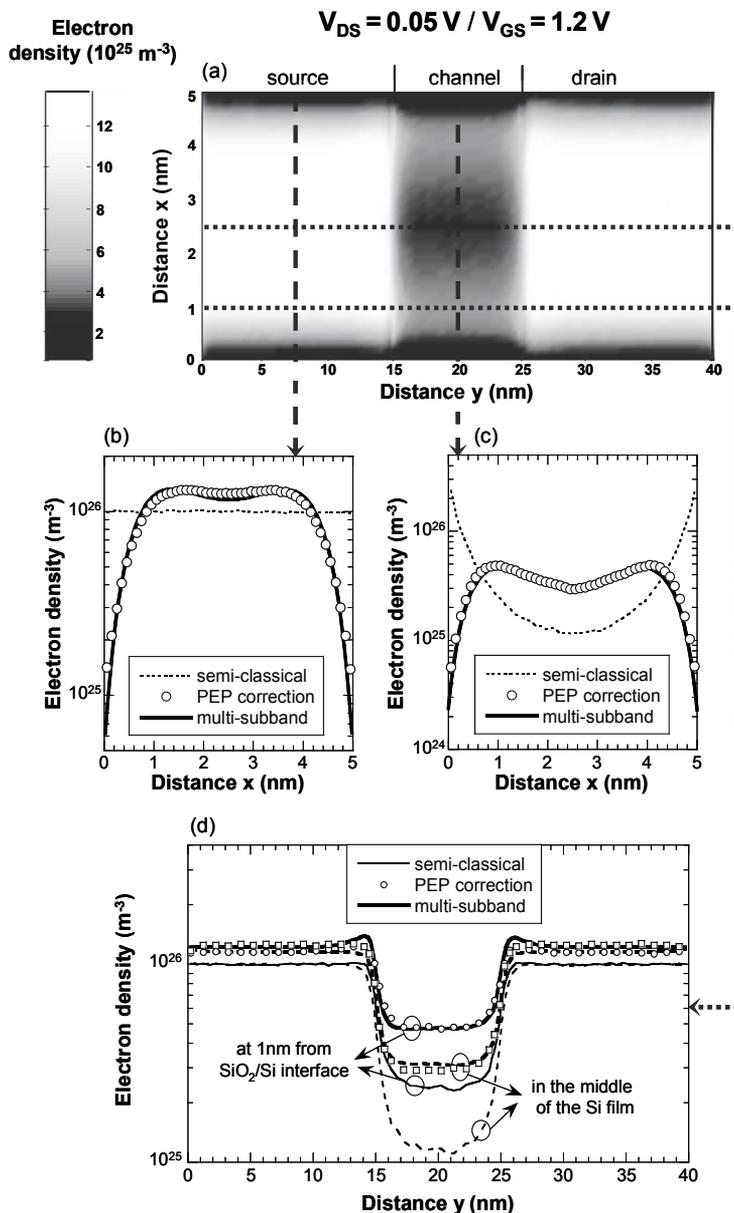


Fig. 12. Cartography of the electron density resulting from the PEP correction at  $V_{GS} = 1.2$  V and  $V_{DS} = 0.05$  V (a). Electron density profiles extracted in different slices of the device along either gate-to-gate (b-c) or source-drain (d) directions for semi-classical (thin lines), PEP corrected (symbols) and multi-subband Monte-Carlo (thick lines) simulations.

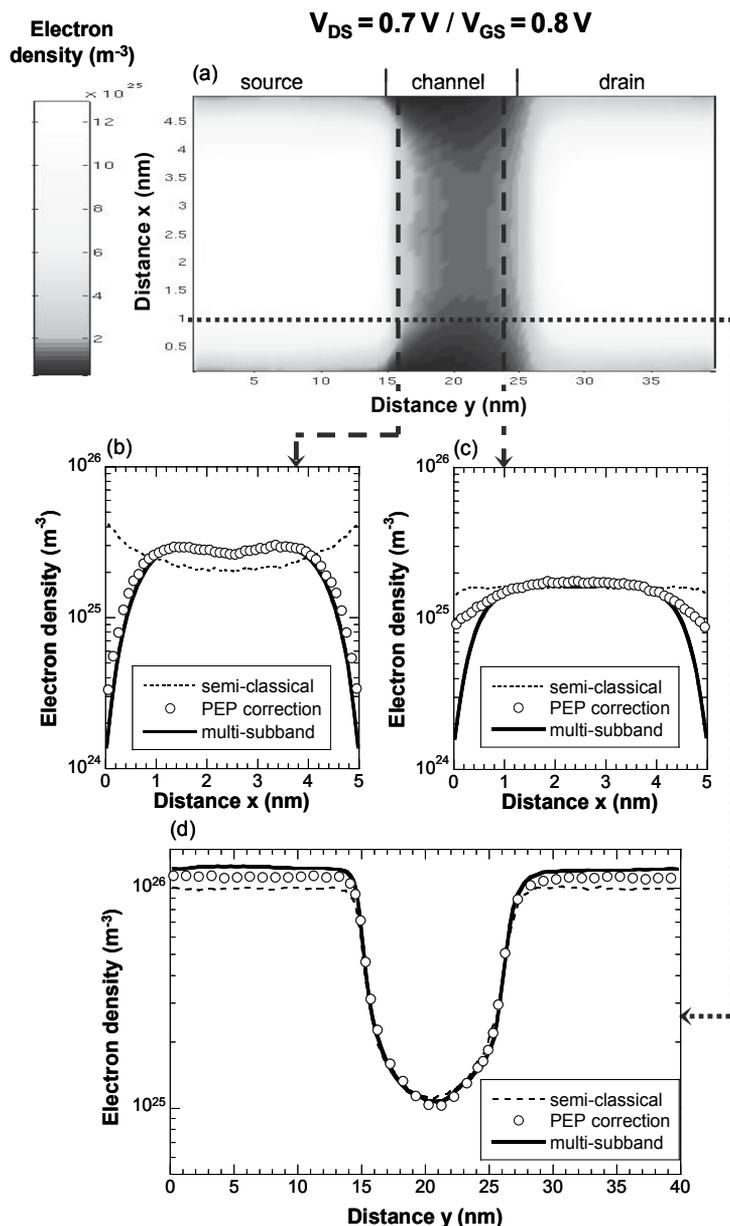


Fig. 13. As in Fig. 12 at  $V_{GS} = 0.8 \text{ V}$  and  $V_{DS} = 0.7 \text{ V}$  for semi-classical (thin dotted lines), PEP corrected (open circles) and multi-subband Monte-Carlo (thick solid lines) simulations.

At these same applied voltages, we also compare electron density and potential profiles resulting from semi-classical, PEP correction and multi-subband Monte-Carlo simulation along the confinement and the transport directions (cf. Figs. 12b-12d ; 13b-13d). First of all, comparisons with semi-classical results highlight the impact of quantum confinement effects. At low drain voltage, two maxima of density at about 1 nm from the Si/SiO<sub>2</sub> interfaces are observed (cf. Figs. 12a-12c). Figs. 13a-13c show a gradual reduction of the

confinement effects along the channel under high drain voltage. It is mainly explained by the curvature of the conduction band that is less sensitive to the drain voltage at the source-end of the channel than at the drain-end. In most cases, excellent agreement is found between PEP and multi-subband electron density profiles all along the device. However, at the drain end of the channel and under high drain voltage (cf. Fig. 13c), the electron repulsion at the Si/SiO<sub>2</sub> interfaces induced by the PEP correction is less pronounced than that induced by MSMC. Such behavior may be related to carrier heating in this channel region, which is observable in Fig. 14a where the electron kinetic energy averaged in the confinement direction is plotted as a function of the source-to-drain distance. Indeed, in MSMC simulation, high electric field induces electron heating which redistributes carriers by phonon scattering within higher subbands with envelope functions different from that of the low-field case. In contrast, using the PEP correction even if the same “quantum” potential barrier is seen by both thermal and hot electrons, hot electrons are allowed to get closer to the oxide interfaces than thermal ones. To still improve the PEP model, an additional correction is probably needed to better describe the repulsive effect for high energy carriers. Besides, kinetic energy resulting from semi-classical and PEP corrected Monte-Carlo are quite similar and higher than that resulting from MSMC. This is consistent with the fact that energy of carriers at thermal equilibrium in a 2D electron gas is  $k_B T_e$  instead of  $3/2 k_B T_e$  in a 3D electron gas (with  $k_B$  the Boltzmann constant and  $T_e$  the electron temperature). However, it should be nicely observed that this error does not really affect the inversion charge. This conservation combined with the additional excellent agreement obtained on the average velocity profile plotted in Fig. 14b consistently explains the good concordance on drain currents calculated with both approaches (cf. Fig. 10).

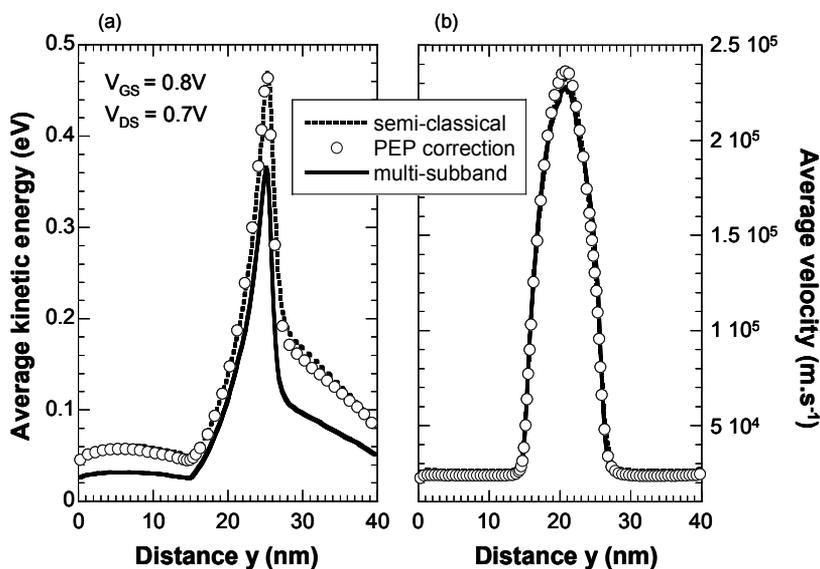


Fig. 14. Kinetic energy (a) and velocity (b) averaged over the confinement direction (semi-classical and PEP correction) or over the different subbands (multi-subband) according to carrier density at  $V_{GS} = 0.8\text{ V}$  and  $V_{DS} = 0.7\text{ V}$  resulting from semi-classical (dotted lines), PEP corrected (open circles) and multi-subband (solid lines) Monte-Carlo simulations.

Moreover, velocity obtained from PEP and multi-subband Monte-Carlo models are similar to the semi-classical one. Contrary to the electrostatics in the confinement direction, the transport properties in the source-to-drain direction are not significantly affected by quantum confinement effects in the case of very thin silicon film.

Finally, the very good overall agreement on both electrical characteristics and microscopic quantities between PEP corrected and multi-subband Monte-Carlo approaches for a very aggressive double-gate MOSFET largely validates the PEP correction. Similar results have been obtained on a double-gate nMOSFET with a channel length  $L_C = 20$  nm and a silicon thickness  $T_{Si} = 8$  nm (Jaud et al, 2007b). By means of comparisons with semi-classical Monte-Carlo results, the PEP correction can now be used to further study the impact of quantum confinement effect on electron transport and device performances depending on design parameters.

## 8. Impact of the quantum confinement effects

In this section, the simulated devices are double-gate nMOSFETs with a silicon film thickness  $T_{Si} = 5$  nm and a channel length  $L_C$  varying from 10 up to 40 nm and double-gate nMOSFETs with  $L_C = 20$  nm and  $T_{Si}$  varying from 5 up to 10 nm. The drive current  $I_{DS}$  resulting from semi-classical and PEP corrected Monte-Carlo simulations and extracted at  $V_{GS} - V_{th} = V_{DS} = V_{DD} = 0.7$  V is plotted in Fig. 15 as a function of  $L_C$  and  $T_{Si}$  where  $V_{th}$  is the threshold voltage obtained at low  $V_{DS}$  for each device.

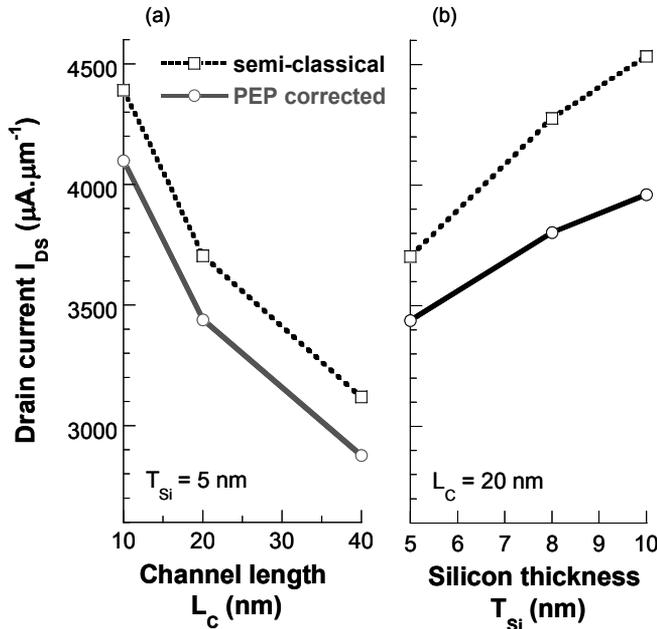


Fig. 15. Drain current  $I_{DS}$  resulting from semi-classical (dotted lines) and PEP corrected (full lines) Monte-Carlo simulations and obtained at  $V_{GS} - V_{th} = V_{DS} = V_{DD} = 0.7$  V on double-gate nMOSFET as a function of the channel length  $L_C$  ( $T_{Si} = 5$  nm) (a) and the silicon film thickness  $T_{Si}$  ( $L_C = 20$  nm) (b).

At these same bias, we plot the inversion charge  $N_{inv}$  and the velocity  $v_{inj}$  extracted at the maximum of the potential energy (averaged in the confinement direction according to carrier density) obtained on the previous devices with  $T_{Si} = 5$  nm (cf. Fig. 16) and  $L_C = 20$  nm (cf. Fig. 17) for semi-classical and PEP corrected Monte-Carlo as a function of the drive current  $I_{DS}$ . For all simulated devices, quantum confinement effects induce a decrease of drive current lower than 13% compared to semi-classical results (cf. Fig. 15), which is mainly attributed to the decrease of the inversion charge. When the channel length is decreased we observe a raise of the drive current  $I_{DS}$  (cf. Fig. 15a) that presents a linear correlation with the increase of the average velocity at the top of the barrier while the inversion charge  $N_{inv}$  is nearly independent on  $L_C$  (cf. Fig. 16). This increase of injection velocity is directly related to the reduced backscattering in the channel.

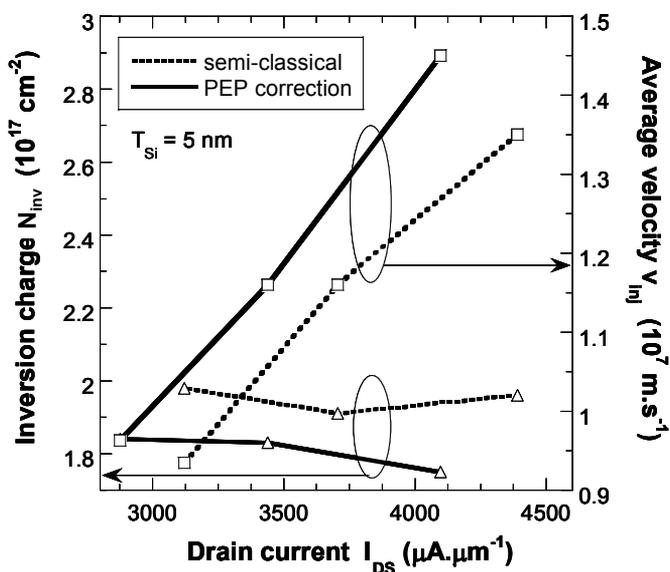


Fig. 16. Inversion charge (triangles) and average velocity (squares) as a function of the drain current  $I_{DS}$  resulting from semi-classical (dotted lines) and PEP corrected (full lines) Monte-Carlo simulations and obtained at  $V_{GS} - V_{th} = V_{DS} = V_{DD} = 0.7$  V on double-gate nMOSFET with different channel lengths  $L_C$  at given  $T_{Si} = 5$  nm. The inversion charge and the average velocity have been extracted at the maximum of the potential energy (averaged over the confinement direction according to carrier density).

When the silicon film thickness is increased we observe an increase of the drive current  $I_{DS}$  (cf. Fig. 15b), that presents a linear correlation with the increase of the inversion charge at the top of the barrier (cf. Fig. 17). This behavior of the inversion charge is a direct consequence of the thinning-induced enhancement of source-access resistance which weakens the gate control of the channel. Surprisingly enough, when the silicon film thickness is reduced (cf. Fig. 15b), the impact of quantum confinement effects becomes less important and the drop of drive current is about 7% for  $T_{Si} = 5$  nm. The increase of the inversion charge distribution in the middle of the film partly compensates the carrier

depletion near the Si/SiO<sub>2</sub> interfaces, which explains this trend. For 5 nm silicon film thickness, the reduction of drive current due to quantum confinement effects is not very sensitive to the channel length  $L_C$  (cf. Fig. 15a); this drive current reduction only slightly increases when  $L_C$  increases.

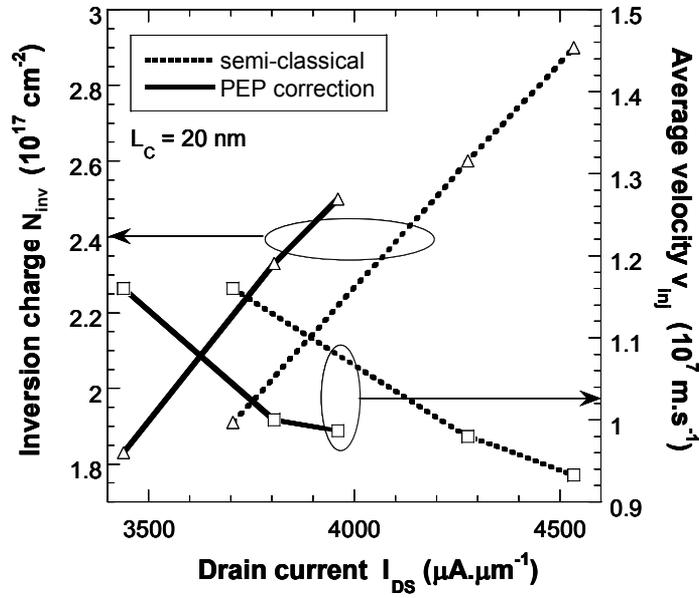


Fig. 17. As in Fig. 16 but obtained on double-gate nMOSFET with different silicon film thicknesses  $T_{Si}$  at given  $L_C = 20 \text{ nm}$ .

## 9. Conclusion

The new Pearson Effective Potential scheme has been developed to account for quantum confinement effects in nano-scaled devices and has been implemented into a semi-classical Monte-Carlo simulator. It mainly consists of an improvement of the particle wave-packet description: the Gaussian distribution used in the usual GEP correction is replaced by a Pearson IV distribution that can much better fit the square modulus of the ground subband Schrödinger wave function. Thanks to a judicious calibration of Pearson IV parameters dependent on the silicon film thickness and local electric field in the confinement direction, the PEP correction accurately predicts electrostatic quantum confinement effects in ultimate bulk, SOI or double-gate nMOS and properly describes the impact of quantum confinement on electron transport in terms of both electrical characteristics and microscopic quantities. Indeed, excellent agreement between quantum corrected and multi-subband Monte-Carlo simulations are shown on a nano-scaled double-gate nMOSFET. Comparisons between semi-classical and PEP corrected Monte-Carlo simulations on nano-scaled double-gate nMOSFETs show that the reduction of the inversion charge induced by quantum confinement effects are mainly responsible for a decrease of about 10% on the drive current.

Finally, the PEP correction can be easily extended to several confinement directions and is well suited for the simulation of various nMOSFET architectures such as double-gate, silicon on insulator or bulk.

## 10. Acknowledgements

This work was supported by the Agence Nationale pour la Recherche through project MODERN (ANR-05-NANO-002).

## 11. Appendix: Pearson IV definition and PEP calibration

The Pearson IV distribution is defined as (Selberherr, 1984 ; Sze, 1988):

$$f(x) = K \left[ b_0 + b_1(x - R_p) + b_2(x - R_p)^2 \right]^{1/2b_2} \exp \left[ -\frac{\frac{b_1}{b_2} + 2b_1}{\sqrt{4b_0b_2 - b_1^2}} a \tan \left( \frac{2b_2(x - R_p) + b_1}{\sqrt{4b_0b_2 - b_1^2}} \right) \right] \quad (3)$$

with  $b_0$ ,  $b_1$  and  $b_2$  given by:

$$b_0 = -\frac{\sigma_p^2 (4\beta - 3\gamma^2)}{10\beta - 12\gamma^2 - 18}$$

$$b_1 = -\frac{\gamma \sigma_p (\beta + 3)}{10\beta - 12\gamma^2 - 18} \quad (4)$$

$$b_2 = -\frac{2\beta - 3\gamma^2 - 6}{10\beta - 12\gamma^2 - 18}$$

and  $K$  is a constant to ensure that the Pearson IV is normalized.

The skewness  $\gamma$  and the kurtosis  $\beta$  obey the following conditions:

$$0 < \gamma^2 < 32 \quad (5)$$

$$\beta > \frac{39\gamma^2 + 48 + 6(\gamma^2 + 4)^{\frac{3}{2}}}{32 - \gamma^2} \quad (6)$$

We recall that the average position  $R_p$ , the standard deviation  $\sigma_p$ , the skewness  $\gamma$  and the kurtosis  $\beta$  are defined as a function of the first four moments of the distribution function as following:

$$R_p = \mu_1 \quad \sigma_p = \sqrt{\mu_2} \quad \gamma = \frac{\mu_3}{\mu_2^{3/2}} \quad \beta = \frac{\mu_4}{\mu_2^2} \quad (7)$$

In our PEP correction, the wave-packet of a particle located in “x” in the confinement direction and under an electric field  $E_x$  is represented by a Pearson IV distribution whose moments have been calibrated as a function of  $E_x$  and  $T_{Si}$ . We present here the expressions of each of the four calibrated Pearson IV moments. Table 2 gathers all the notations specifying their unit and significance. The parameters’ values necessary for Pearson moments calculation are listed in Table 3.

Name	Unit	Definition
$\alpha_1$	m <sup>-1</sup>	Constant parameter for $\sigma_p$ calculation $\alpha_1=10^9$ m <sup>-1</sup>
$\alpha_2$	m <sup>-1</sup>	Constant parameter for $\sigma_p$ calculation $\alpha_2=17.10^{11}$ m <sup>-1</sup>
$\beta$		Kurtosis (cf. eq. 12)
$ E_x $	V.m <sup>-1</sup>	Local electric field in the confinement direction
$ E_x _{\max}$	V.m <sup>-1</sup>	Constant parameter $ E_x _{\max}= 3.5 \cdot 10^8$ V.m <sup>-1</sup>
$\gamma$	ad.	Skewness (cf. eq. 11)
$\gamma_{\max}$	ad.	Parameter for $\gamma$ calculation (cf. Table 3)
$R_p$	m	Average position (cf. eq. 9)
$R_{Pa}$	ad.	Parameter for $R_p$ calculation (cf. Table 3)
$R_{Pdiv}$	m	Parameter for $R_p$ calculation (cf. Table 3)
$R_{Pmax}$	ad.	Parameter for $R_p$ calculation (cf. Table 3)
$R_{p0}$	m	Average position of a carrier under a zero electric field $E_x$
$R_{p1}$	m	Average position of a carrier located at the 1 <sup>st</sup> interface (cf. eq. 8)
$R_{p2}$	m	Average position of a carrier located at the 2 <sup>nd</sup> interface (cf. eq. 8)
$\sigma_p$	m	Standard deviation (cf. eq. 10)
$T_{Si}$	m	Silicon film thickness
$T_{Sis}$	ad.	Parameter for $\sigma_p$ calculation (cf. Table 3)
x1	m	Location of the 1 <sup>st</sup> interface
x2	m	Location of the 2 <sup>nd</sup> interface

Table 2. Unit and significance of all the notations used for the calculation of the Pearson IV calibrated parameters (ad. is for adimensional).

### Average position

The average position is calculated in two different steps. Firstly, the average position of a particle located at the first interface ( $R_{P1}$ ) and at the second interface ( $R_{P2}$ ) are calculated as a function of  $E_x$  and  $T_{Si}$  so as to fit the theoretical values:

$$R_P = \frac{T_{Si}}{2} - \frac{1}{\log(10^{R_{Pa}})} \left( \frac{T_{Si}}{2} - R_{P_{max}} \right) \times \log \left[ \frac{10^{R_{Pa}} \times |E_x|}{|E_x|_{max}} + 1 \right] \quad (8)$$

Moreover, for a particle under a zero electric field, the average position of its wave-packet ( $R_{P0}$ ) is equal to its location. Secondly, for each particle location, the average position of its wave-packet  $R_P$  is calculated from  $R_{P0}$ ,  $R_{P1}$  and  $R_{P2}$  while ensuring that  $R_P(x)$  is continuous and regular:

$$\begin{aligned} &\text{If } x \leq R_{P0} \\ &\text{then } R_P = R_{P0} + (R_{P0} - R_{P1} - x_1) \cdot \frac{\tanh\left(\frac{(x - R_{P0})}{R_{Pdiv}}\right)}{\left| \tanh\left(\frac{(x_1 - R_{P0})}{R_{Pdiv}}\right) \right|} \\ &\text{else } R_P = R_{P0} - (R_{P0} + R_{P2} - x_2) \cdot \frac{\tanh\left(\frac{(x - R_{P0})}{R_{Pdiv}}\right)}{\left| \tanh\left(\frac{(x_2 - R_{P0})}{R_{Pdiv}}\right) \right|} \end{aligned} \quad (9)$$

### Standard deviation

For a particle under a local electric field in the confinement direction  $E_x$ , the standard deviation of the Pearson IV representing its wave-packet is calculated as follows:

$$\sigma_P = \frac{1}{\alpha_1} \times \left[ \log(T_{Sis}) + \frac{T_{Sis} + 1.5}{50} \right] - \frac{1}{\alpha_2} \times (T_{Sis})^3 \times \log \left[ \frac{(-8 \times T_{Sis} + 90) \times |E_x|}{|E_x|_{max}} + 1 \right] \quad (10)$$

### Skewness

The skewness is calculated as a function of  $E_x$  and  $T_{Si}$  so as to fit the theoretical values:

$$\gamma = \gamma_{max} \times \tanh \left[ \frac{T_{Si}}{10^{-9}} \times \frac{|E_x|}{|E_x|_{max}} \right] \quad (11)$$

Moreover, the sign of the skewness is then adjust to be in adequacy with the sign of the local electric field in the confinement direction  $E_x$ .

### Kurtosis

In accordance with Pearson IV definition (Selberherr, 1984 ; Sze, 1988), the kurtosis is only calculated as a function of the skewness  $\gamma$  so as to be minimal and closest to the Gaussian value:

$$\beta = \frac{39\gamma^2 + 48 + 6(\gamma^2 + 4)^{\frac{3}{2}}}{32 - \gamma^2} + \varepsilon \quad (12)$$

with  $\varepsilon > 0$  to prevent from numerical difficulties.

Name	$T_{Si} < 10 \text{ nm}$	$T_{Si} \geq 10 \text{ nm}$
$\gamma_{\max}$	$0.03 \times T_{Si} / 10^{-9} + 0.6$	0.9
$R_{Pa}$	5	Integer part $[0.7 \times T_{Si} / 10^{-9} - 2]$
$R_{Pdiv}$	$6 \times 10^{-9}$	$0.4 \times T_{Si} + 2 \times 10^{-9}$
$R_{Pmax}$	$-0.034 \times T_{Si} + 1.17 \times 10^{-9}$	$0.83 \times 10^{-9}$
$T_{Sis}$	$T_{Si} / 10^{-9}$	10

Table 3. Values of the parameters as a function of  $T_{Si}$  used for Pearson IV calibrated parameters calculation according to the units defined in Table 2.

## 12. References

- Ahmed et al., 2005 : S.S. Ahmed, C. Ringhofer, D. Vasileska, Parameter-Free Effective Potential Method for Use in Particle-Based Device Simulations, *IEEE Transactions on Nanotechnology*, vol. 4, pp. 465-471, 2005.
- Akis et al., 2001 : R. Akis, N. Milicic, D. K. Ferry, D. Vasileska, An effective potential method for including quantum effects into the simulation of ultra-short and ultra-narrow channel MOSFETs, *International conference on Modeling and Simulation of Microsystem (MSM 2001)*, vol. 1, pp. 550-553, 2001.
- Bohm, 1952: D. Bohm, A suggested interpretation of the quantum theory in terms of "Hidden" variables. II, *Physical Review*, vol. 85, pp. 180-193, 1952.
- Dhatt et al., 2005 : G. Dhatt, G. Touzot, E. Lefrançois, Méthode des éléments finis, *Hermès Science Publications*, pp. 345-383, 2005.
- Fan et al., 2004 : X.-F. Fan, X. Wang, B. Winstead, L. F. Register, U. Ravaioli, S. Banerjee, MC simulations of Strained-Si MOSFET with Full-Band Structure and Quantum Correction, *IEEE Transaction on Electron Devices*, vol. 51, pp. 962-970, 2004.
- Ferry et al., 2000 : D.K. Ferry, R. Akis, D. Vasileska, Quantum effects in MOSFETs: Use of an Effective Potential in 3D Monte Carlo Simulation of Ultra-Short Channel Devices, *IEDM Technical Digest*, pp. 287-290, 2000.
- Feynman & Hibbs, 1965 : R.P. Feynman and A. R. Hibbs, Quantum mechanics and path integrals, *McGraw-Hill Publishing Company*, pp. 267-286, 1965.
- Jaud et al., 2006 : M.-A. Jaud, S. Barraud, P. Dollfus, H. Jaouen, F. de Crecy and G. Le Carval, Validity of the effective potential approach for the simulation of quantum confinement effects: A Monte-Carlo study, *Journal of Comput. Electron.*, vol. 5, pp. 171-175, 2006.
- Jaud et al., 2007a : M.-A. Jaud, S. Barraud, P. Dollfus, H. Jaouen and G. Le Carval, Pearson versus gaussian effective potentials for quantum-corrected Monte-Carlo simulation, *Journal of Comput. Electron.*, vol. 6, pp. 19-22, 2007.

- Jaud et al., 2007b : M.-A. Jaud, S. Barraud, J. Saint-Martin, A. Bournel, P. Dollfus and H. Jaouen, Pearson Effective Potential vs. Multi-Subband Monte-Carlo Simulation for Electron Transport in DG nMOSFET, in *Proc. SISPAD 2007*, Springer, pp. 65-68, 2007.
- Jaud et al., 2008 : M.-A. Jaud, S. Barraud, J. Saint-Martin, A. Bournel, P. Dollfus and H. Jaouen, A Pearson Effective Potential for Monte-Carlo simulation of quantum confinement effects in nMOSFETs, *IEEE Transactions on Electron Devices*, vol. 55, pp. 3450-3458, 2008.
- Li et al., 2002 : Y. Li, T.-W. Tang and X. Wang, Modeling of quantum effects for ultrathin oxide MOS structures with an effective potential, *IEEE Transactions on Nanotechnology*, vol. 1, pp. 238-242, 2002.
- Lucci et al., 2005 : L. Lucci, P. Palestri, D. Esseni, L. Selmi, Multi-subband Monte Carlo modeling of nano-MOSFETs with strong vertical quantization and electron gas degeneration, *IEDM Tech. Dig.*, pp. 617-620, 2005.
- Madelung, 1926 : E. Madelung, Quantatheorie in hydrodynamischer form, *Z. Phys.*, vol. 40, pp. 322-326, 1926
- Nedjalkov et al., 2004 : M. Nedjalkov, H. Kosina, S. Selberherr, C. Ringhofer, D. K. Ferry, Unified particle approach to Wigner-Boltzmann transport in small semiconductor devices, *Physical Review B*, vol. 70, p. 115319, 2004.
- Palestri et al., 2005 : P. Palestri, S. Eminent, D. Esseni, C. Fiegna, E. Sangiorgi, L. Selmi, An improved semi-classical Monte-Carlo approach for nano-scale MOSFET simulation, *Solid-State Electronics*, vol. 49, pp. 727-732, 2005.
- Querlioz et al., 2006 : D. Querlioz, P. Dollfus, V.-N. Do, A. Bournel and V. Lien Nguyen, An improved Wigner Monte-Carlo technique for the self-consistent simulation of RTDs, *Journal of Computational Electronics*, vol. 5 pp. 443-446, 2006.
- Querlioz et al., 2007 : D. Querlioz, J. Saint-Martin, K. Huet, A. Bournel, V. Aubry-Fortuna, C. Chassat, S. Galdin-Retailleau and P. Dollfus, On the ability of the particle Monte Carlo technique to include quantum effects in nano-MOSFET simulation, *IEEE Transaction on Electron Devices*, vol. 54, pp. 2232-2242, 2007.
- Riolino et al., 2006 : I. Riolino, M. Braccioli, L. Lucci, D. Esseni, C. Fiegna, P. Palestri and L. Selmi, Monte-Carlo simulation of decananometric double-gate SOI devices: Multi-subband vs. 3D electron gas with quantum corrections, *ESSDERC Tech. Digest.*, pp. 162-165, 2006.
- Saint Martin et al., 2004 : J. Saint-Martin, A. Bournel and P. Dollfus, On the ballistic transport in nanometer scaled DG MOSFET, *IEEE Transactions on Electron Devices*, vol. 51, pp. 1148-1155, 2004.
- Saint Martin et al., 2006 : J. Saint-Martin, A. Bournel, F. Monsef, C. Chassat and P. Dollfus, Multi sub-band Monte Carlo simulation of an ultra-thin double-gate MOSFET with 2D electron gas, *Semiconductor Science and Technology*, vol. 21, L29-L31, 2006.
- Selberherr, 1984 : S. Selberherr, Analysis and simulation of semiconductor devices, Springer-Verlag Wien New-York, p. 46 and following, 1984.
- Shifren et al., 2003 : L. Shifren, C. Ringhofer and D. K. Ferry, A Wigner Function-Based Quantum Ensemble Monte-Carlo study of a Resonant Tunneling Diode, *IEEE Transactions on Electron Devices*, vol. 50, pp. 769-773, 2003.

- Sverdlov et al., 2005 : V. Sverdlov, A. Gehring, H. Kosina and S. Selberherr, Quantum transport in ultra-scaled double-gate MOSFETs: A Wigner function based Monte-Carlo approach, *Solid-State Electronics*, vol. 49, pp. 1510-1515, 2005.
- Sze et al., 1988 : S. M. Sze, VLSI Technology, second edition, *McGraw-Hill Book Company*, p. 332 and following, 1988.
- Tang et al., 2003 : T. Tang and B. Wu, Quantum corrected Monte Carlo simulation of semiconductor devices using the effective conduction-band edge method, *Journal of Computational Electronics*, vol. 2, pp. 131-135, 2003.
- Tsuchiya et al., 2003 : H. Tsuchiya, M. Horino and T. Miyoshi, Quantum Monte Carlo device simulation of nano-scaled SOI-MOSFETs, *Journal of Computational Electronics*, vol. 2, pp. 91-95, 2003.

# Monte Carlo Device Simulations

Dragica Vasileska<sup>1</sup>, Katerina Raleva<sup>2</sup> and Stephen M. Goodnick<sup>1</sup>

<sup>1</sup>*Arizona State University, Tempe AZ*

<sup>2</sup>*University Sts Cyril and Methodius, Skopje,*

<sup>1</sup>*USA*

<sup>2</sup>*Republic of Macedonia*

## 1. Introduction

As semiconductor devices are scaled into nanoscale regime, first velocity saturation starts to limit the carrier mobility due to pronounced intervalley scattering, and when the device dimensions are scaled to 100 nm and below, velocity overshoot starts to dominate the device behavior leading to larger ON-state currents. Alongside with the developments in the semiconductor nanotechnology, in recent years there has been significant progress in physical based modeling of semiconductor devices. First, for devices for which gradual channel approximation can not be used due to the two-dimensional nature of the electrostatic potential and the electric fields driving the carriers from source to drain, drift-diffusion models have been exploited. These models are valid, in general, for large devices in which the fields are not that high so that there is no degradation of the mobility due to the electric field. The validity of the drift-diffusion models can be extended to take into account the velocity saturation effect with the introduction of field-dependent mobility and diffusion coefficients. When velocity overshoot becomes important, drift diffusion model is no longer valid and hydrodynamic model must be used. The hydrodynamic model has been the workhorse for technology development and several high-end commercial device simulators have appeared including Silvaco, Synopsys, Crosslight, etc. The advantages of the hydrodynamic model are that it allows quick simulation runs but the problem is that the amount of the velocity overshoot depends upon the choice of the energy relaxation time. The smaller is the device, the larger is the deviation when using the same set of energy relaxation times. A standard way in calculating the energy relaxation times is to use bulk Monte Carlo simulations. However, the energy relaxation times are material, device geometry and doping dependent parameters, so their determination ahead of time is not possible. To avoid the problem of the proper choice of the energy relaxation times, a direct solution of the Boltzmann Transport Equation (BTE) using the Monte Carlo method is the best method of choice. That is why the focus of this review paper is on explaining basic Monte Carlo device simulator and then the focus will be shifted on the inclusion of various higher order effects that explain particular physical phenomena or processes.

The Monte Carlo book chapter is organized as follows. First, the idea behind the Monte Carlo technique is outlined by revoking the path integral method for the solution of the BTE. This approach naturally leads to the free-flight-scatter sequence that is used in solving the BTE using the Monte Carlo method. Various scattering mechanisms relevant for different

materials are given to completely specify the collision integral in the BTE. A discussion followed with the presentation of a generic flow-chart for implementing bulk Monte Carlo code is presented. Note that bulk Monte Carlo approach is suitable for the characterization of materials, but in order to study behavior of semiconductor devices coupling of the Monte Carlo transport kernel with a Poisson equation solver which gives the self-consistent field that moves the carriers around is needed. Important ingredients in describing particle-based device simulators are the particle-mesh coupling, treatment of the Ohmic contacts and calculation of the current. A generic flowchart of a particle-based device simulator is provided. The prospects of the Monte Carlo method for the solution of the Boltzmann transport equation, in the context of device simulations of nanoscale structures and of solar cells and power devices, are discussed at the end of the book chapter.

## 2. Importance of MC particle-based device simulations

### 2.1 Industry trends and the need for modeling and simulation

As semiconductor feature sizes shrink into the nanometer scale regime, even conventional device behavior becomes increasingly complicated as new physical phenomena at short dimensions occur, and limitations in material properties are reached [1]. In addition to the problems related to the understanding of actual operation of ultra-small devices, the reduced feature sizes require more complicated and time-consuming manufacturing processes. This fact signifies that a pure trial-and-error approach to device optimization will become impossible since it is both too time consuming and too expensive. Since computers are considerably cheaper resources, simulation is becoming an indispensable tool for the device engineer. Besides offering the possibility to test hypothetical devices which have not (or could not) yet been manufactured, simulation offers unique insight into device behavior by allowing the observation of phenomena that can not be measured on real devices. *Computational Electronics* [2,3,4] in this context refers to the physical simulation of semiconductor devices in terms of charge transport and the corresponding electrical behavior. It is related to, but usually separate from process simulation, which deals with various physical processes such as material growth, oxidation, impurity diffusion, etching, and metal deposition inherent in device fabrication [5] leading to integrated circuits. Device simulation can be thought of as one component of technology for computer-aided design (TCAD), which provides a basis for device modeling, which deals with compact behavioral models for devices and sub-circuits relevant for circuit simulation in commercial packages such as SPICE [6]. The relationship between various simulation design steps that have to be followed to achieve certain customer need is illustrated in Figure 1.

The goal of *Computational Electronics* is to provide simulation tools with the necessary level of sophistication to capture the essential physics while at the same time minimizing the computational burden so that results may be obtained within a reasonable time frame. Figure 2 illustrates the main components of semiconductor device simulation at any level. There are two main kernels, which must be solved self-consistently with one another, the transport equations governing charge flow, and the fields driving charge flow. Both are coupled strongly to one another, and hence must be solved simultaneously. The fields arise from external sources, as well as the charge and current densities which act as sources for the time varying electric and magnetic fields obtained from the solution of Maxwell's equations. Under appropriate conditions, only the quasi-static electric fields arising from the solution of Poisson's equation are necessary.

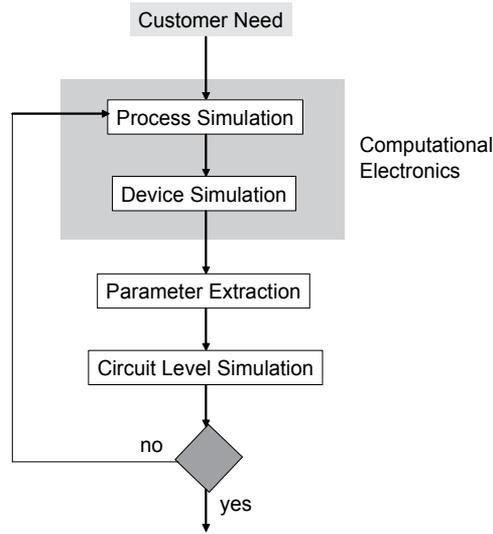


Fig. 1. Design sequence to achieve desired customer need.

The fields, in turn, are driving forces for charge transport as illustrated in Figure 3 for the various levels of approximation within a hierarchical structure ranging from compact modeling at the top to an exact quantum mechanical description at the bottom. At the very beginnings of semiconductor technology, the electrical device characteristics could be estimated using simple analytical models (gradual channel approximation for MOSFETs) relying on the drift-diffusion (DD) formalism. Various approximations had to be made to obtain closed-form solutions, but the resulting models captured the basic features of the devices [7]. These approximations include simplified doping profiles and device geometries. With the ongoing refinements and improvements in technology, these approximations lost their basis and a more accurate description was required. This goal could be achieved by solving the DD equations numerically. Numerical simulation of carrier transport in semiconductor devices, dates back to the famous work of Scharfetter and Gummel [8], who proposed a robust discretization of the DD equations, which is still in use today.

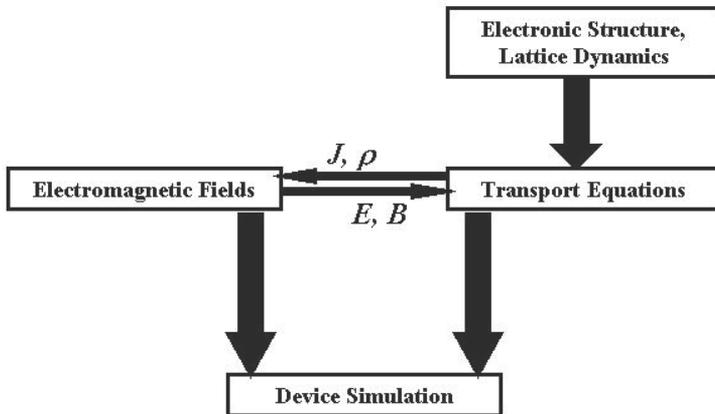


Fig. 2. Schematic description of the device simulation sequence.

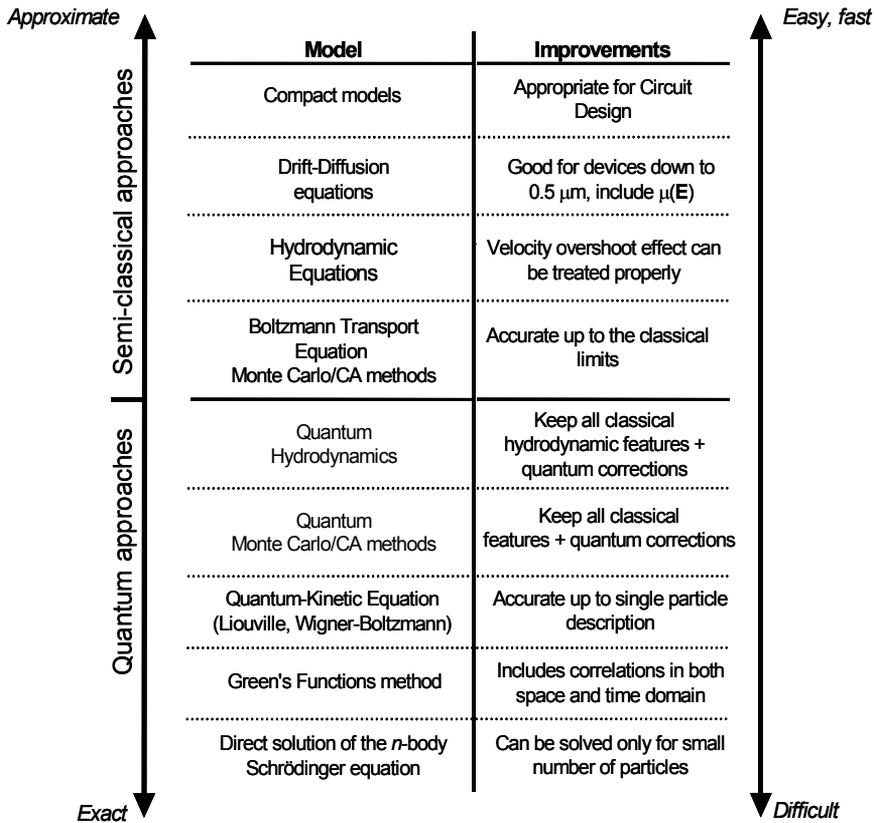


Fig. 3. Illustration of the hierarchy of transport models.

However, as semiconductor devices were scaled into the submicrometer regime, the assumptions underlying the DD model lost their validity. Therefore, the transport models have been continuously refined and extended to more accurately capture transport phenomena occurring in these devices. The need for refinement and extension is primarily caused by the ongoing feature size reduction in state-of-the-art technology. As the supply voltages can not be scaled accordingly without jeopardizing the circuit performance, the electric field inside the devices has increased. A large electric field, which rapidly changes over small length scales, gives rise to non-local and hot-carrier effects which begin to dominate device performance. An accurate description of these phenomena is required and is becoming a primary concern for industrial applications.

To overcome some of the limitations of the DD model, extensions have been proposed which basically add an additional balance equation for the average carrier energy [9]. Furthermore, an additional driving term is added to the current expression which is proportional to the gradient of the carrier temperature. However, a vast number of these models exist, and there is a considerable amount of confusion as to their relation to each other. It is now a common practice in industry to use standard hydrodynamic models in trying to understand the operation of as-fabricated devices, by adjusting any number of phenomenological parameters (e.g. mobility, impact ionization coefficient, etc.). However, such tools do not have predictive capability for ultra-small structures, for which it is

necessary to relax some of the approximations in the Boltzmann transport equation [10]. Therefore, one needs to move downward to the quantum transport area in the hierarchical map of transport models shown in Figure 3, where, at the very bottom we have the Green's function approach [11,12,13]. The latter is the most exact, but at the same time the most difficult of all. In contrast to, for example, the Wigner function approach (which is Markovian in time), the Green's functions method allows one to consider simultaneously correlations in space and time, both of which are expected to be important in nano-scale devices. However, the difficulties in understanding the various terms in the resultant equations and the enormous computational burden needed for its actual implementation make the usefulness in understanding quantum effects in actual devices of limited values. For example, the only successful utilization of the Green's function approach commercially is the NEMO (Nano-Electronics Modeling) simulator [14], which is effectively 1D and is primarily applicable to resonant tunneling diodes.

From the discussion above it follows that, contrary to the recent technological advances, the present state of the art in device simulation is currently lacking in the ability to treat these new challenges in scaling of device dimensions from conventional down to quantum scale devices. For silicon devices with active regions below 0.2 microns in diameter, macroscopic transport descriptions based on drift-diffusion models are clearly inadequate. As already noted, even standard hydrodynamic models do not usually provide a sufficiently accurate description since they neglect significant contributions from the tail of the phase space distribution function in the channel regions [15,16]. Within the requirement of self-consistently solving the coupled transport-field problem in this emerging domain of device physics, there are several computational challenges, which limit this ability. One is the necessity to solve both the transport and the Poisson's equations over the full 3D domain of the device (and beyond if one includes radiation effects). As a result, highly efficient algorithms targeted to high-end computational platforms (most likely in a multi-processor environment) are required to fully solve even the appropriate field problems. The appropriate level of approximation necessary to capture the proper non-equilibrium transport physics, relevant to a future device model, is an even more challenging problem both computationally and from a fundamental physics framework.

## 2.2 Drift-Diffusion and hydrodynamic models

In Section 1.1 above, we discussed the various levels of approximations that are employed in the modeling of semiconductor devices. The direct solution of the full BTE is challenging computationally, particularly when combined with field solvers for device simulation. Therefore, for traditional semiconductor device modeling, the predominant model corresponds to solutions of the so-called drift-diffusion equations, which are 'local' in terms of the driving forces (electric fields and spatial gradients in the carrier density), i.e. the current at a particular point in space only depends on the instantaneous electric fields and concentration gradient at that point. The complete drift-diffusion model is based on the following set of equations:

Current equations:

$$\begin{aligned} J_n &= qn(x)\mu_n E(x) + qD_n \frac{dn}{dx} \\ J_p &= qp(x)\mu_p E(x) - qD_p \frac{dp}{dx} \end{aligned} \quad (1)$$

Continuity equations:

$$\begin{aligned}\frac{\partial n}{\partial t} &= \frac{1}{q} \nabla \cdot \mathbf{J}_n + U_n \\ \frac{\partial p}{\partial t} &= -\frac{1}{q} \nabla \cdot \mathbf{J}_p + U_p\end{aligned}\quad (2)$$

Poisson's equation:

$$\nabla \cdot (\epsilon \nabla V) = -(p - n + N_D^+ - N_A^-), \quad (3)$$

where  $U_n$  and  $U_p$  are the net generation-recombination rates .

The continuity equations are the conservation laws for the carriers. A numerical scheme which solves the continuity equations should

- Conserve the total number of particles inside the device being simulated.
- Respect local positive definite nature of carrier density. Negative density is unphysical.
- Respect monotonicity of the solution (i.e. it should not introduce spurious space oscillations).

Conservative schemes are usually achieved by subdivision of the computational domain into patches (boxes) surrounding the mesh points. The currents are then defined on the boundaries of these elements, thus enforcing conservation (the current exiting one element side is exactly equal to the current entering the neighboring element through the side in common). In the absence of generation-recombination terms, the only contributions to the overall device current arise from the contacts. Remember that, since electrons have negative charge, the particle flux is opposite to the current flux. When the equations are discretized, using finite differences for instance, there are limitations on the choice of mesh size and time step [17]:

- The mesh size  $\Delta x$  is limited by the Debye length.
- The time step is limited by the dielectric relaxation time.

A mesh size must be smaller than the Debye length where one has to resolve charge variations in space. A simple example is the carrier redistribution at an interface between two regions with different doping levels. Carriers diffuse into the lower doped region creating excess carrier distribution which at equilibrium decays in space down to the bulk concentration with approximately exponential behavior. The spatial decay constant is the Debye length

$$L_D = \sqrt{\frac{\epsilon k_B T}{q^2 N}} \quad (4)$$

where  $N$  is the doping density,  $\epsilon$  is the dielectric constant,  $k_B$  is the Boltzmann constant,  $T$  is the lattice temperature and  $q$  is the elementary charge. In GaAs and Si, at room temperature the Debye length is approximately 400 Å when  $N \approx 10^{16} \text{ cm}^{-3}$  and decreases to about only 50 Å when  $N \approx 10^{18} \text{ cm}^{-3}$ .

The dielectric relaxation time, on the other hand, is the characteristic time for charge fluctuations to decay under the influence of the field that they produce. The dielectric relaxation time may be estimated using

$$t_{dr} = \frac{\varepsilon}{qN\mu} \quad (5)$$

where  $\mu$  is the carrier mobility.

The drift-diffusion semiconductor equations constitute a coupled nonlinear set. It is not possible, in general, to obtain a solution directly in one step, but a nonlinear iteration method is required. The two most popular methods for solving the discretized equations are the Gummel's iteration method [18] and the Newton's method [19]. It is very difficult to determine an optimum strategy for the solution, since this will depend on a number of details related to the particular device under study.

Finally, the discretization of the continuity equations in conservation form requires the determination of the currents on the mid-points of mesh lines connecting neighboring grid nodes. Since the solutions are accessible only on the grid nodes, interpolation schemes are needed to determine the currents. The approach by Scharfetter and Gummel [8] has provided an optimal solution to this problem, although the mathematical properties of the proposed scheme have been fully recognized much later.

In the computational electronics community, the necessity for the hydrodynamic (HD) transport model is normally checked by comparison of simulation results for HD and DD simulations. Despite the obvious fact that, depending on the equation set, different principal physical effects are taken into account, the influence on the models for the physical parameters is more subtle. The main reason for this is that in the case of the HD model, information about average carrier energy is available in form of carrier temperature. Many parameters depend on this average carrier energy, e.g., the mobilities and the energy relaxation times. In the case of the DD model, the carrier temperatures are assumed to be in equilibrium with the lattice temperature, that is  $T_C = T_L$ , hence, all energy dependent parameters have to be modeled in a different way.

### 2.2.1 Extensions of the Drift-Diffusion model

In the DD approach, the electron gas is assumed to be in thermal equilibrium with the lattice temperature ( $T_n = T_L$ ). However, in the presence of a strong electric field, electrons gain energy from the field and the temperature  $T_n$  of the electron gas is elevated. Since the pressure of the electron gas is proportional to  $nk_B T_n$ , the driving force now becomes the pressure gradient rather than merely the density gradient. This introduces an additional driving force, namely, the temperature gradient besides the electric field and the density gradient. Phenomenologically, one can write the electron current density equation as

$$\mathbf{J} = q(n\mu_n \mathbf{E} + D_n \nabla n + nD_T \nabla T_n) \quad (6)$$

where  $D_T$  is the thermal diffusivity and  $D_n$  is the diffusion constant.

### 2.2.2 Stratton's approach

One of the first derivations of extended transport equations was performed by Stratton [20]. First the distribution function is split into the even and odd parts

$$f(\mathbf{k}, \mathbf{r}) = f_0(\mathbf{k}, \mathbf{r}) + f_1(\mathbf{k}, \mathbf{r}). \quad (7)$$

From  $f_1(-k, r) = -f_1(k, r)$ , it follows that  $\langle f_1 \rangle = 0$ . Assuming that the collision operator  $C$  in the Boltzmann transport equation is linear and invoking the microscopic relaxation time approximation for the collision operator

$$C[f] = -\frac{f - f_{eq}}{\tau(\varepsilon, r)} \quad (8)$$

the BTE can be split into two coupled equations. In particular  $f_1$  is related to  $f_0$  via

$$f_1 = -\tau(\varepsilon, r) \left( \mathbf{v} \cdot \nabla_r f_0 - \frac{q}{\hbar} \mathbf{E} \cdot \nabla_k f_0 \right). \quad (9)$$

The microscopic relaxation time is then expressed by a power law

$$\tau(\varepsilon) = \tau_0 \left( \frac{\varepsilon}{k_B T_L} \right)^{-p}. \quad (10)$$

When  $f_0$  is assumed to be heated Maxwellian distribution, the following equation system is obtained

$$\begin{aligned} \nabla \cdot \mathbf{J} &= q \frac{\partial n}{\partial t} \\ \mathbf{J} &= qn\mu\mathbf{E} + k_B \nabla (n\mu T_n) \\ \nabla \cdot (n\mathbf{S}) &= -\frac{3}{2} k_B \partial (nT_n) + \mathbf{E} \cdot \mathbf{J} - \frac{3}{2} k_B n \frac{T_n - T_L}{\tau_\varepsilon} \\ n\mathbf{S} &= -\left( \frac{5}{2} - p \right) \left( \mu n k_B T_n \mathbf{E} + \frac{k_B^2}{q} \nabla (n\mu T_n) \right) \end{aligned} \quad (11)$$

Equation for the current density can be rewritten as:

$$\mathbf{J} = q\mu \left( n\mathbf{E} + \frac{k_B}{q} T_n \nabla n + \frac{k_B}{q} n(1 + \nu_n) \nabla T_n \right), \quad (12a)$$

with

$$\nu_n = \frac{T_n}{\mu} \frac{\partial \mu}{\partial T_n} = \frac{\partial \ln \mu}{\partial \ln T_n} \quad (12b)$$

which is commonly used as a fit parameter with values in the range [-0.5,-1.0]. For  $\nu_n = -1.0$ , the thermal distribution term disappears. The problem with Eq. (10) for  $\tau$  is that  $p$  must be approximated by an average value to cover the relevant processes. In the particular case of impurity scattering,  $p$  can be in the range [-1.5,0.5], depending on charge screening. Therefore, this average depends on the doping profile and the applied field; thus, no unique

value for  $p$  can be given. Note also that the temperature  $T_n$  is a parameter of the heated Maxwellian distribution, which has been assumed in the derivation. Only for parabolic bands and a Maxwellian distribution, this parameter is equivalent to the normalized second-order moment.

### 2.2.3 Balance equations model

The first three balance equations, derived by taking moments of Boltzmann Transport Equation (BTE), take the form:

$$\begin{aligned} \frac{\partial n}{\partial t} &= \frac{1}{e} \nabla \cdot \mathbf{J}_n + S_n \\ \frac{\partial J_z}{\partial t} &= \frac{2e}{m^*} \sum_i \frac{\partial W_{iz}}{\partial x_i} + \frac{ne^2}{m^*} E_z - \left\langle \left\langle \frac{1}{\tau_m} \right\rangle \right\rangle J_z \\ \frac{\partial W}{\partial t} &= -\nabla \cdot \mathbf{F}_W + \mathbf{E} \cdot \mathbf{J} - \left\langle \left\langle \frac{1}{\tau_E} \right\rangle \right\rangle (W - W_0) \end{aligned} \quad (13)$$

The balance equation for the carrier density introduces the carrier current density, which balance equation introduces the kinetic energy density. The balance equation for the kinetic energy density, on the other hand, introduces the energy flux. Therefore, a new variable appears in the hierarchy of balance equations and the set of infinite balance equations is actually the solution of the BTE. The momentum and energy relaxation rates, that appear in Eq. (13) are ensemble averaged quantities. For simple scattering mechanisms one can utilize the drifted-Maxwellian form of the distribution function, but for cases where several scattering mechanisms are important, one must use bulk Monte Carlo simulations to calculate these quantities.

One can express the energy flux that appears in Eq. (13) in terms of the temperature tensor. The energy flux, is calculated using

$$\mathbf{F}_W = \frac{1}{V} \sum_{\mathbf{p}} \mathbf{v} E(\mathbf{p}) f(\mathbf{r}, \mathbf{p}, t), \quad (14)$$

which means that the  $i$ -th component of this vector equals to

$$F_{Wi} = v_{di} W + nk_B \sum_j T_{ij} v_{dj} + Q_i \quad (15)$$

where  $Q_i$  is the component of the heat flux vector which describes loss of energy due to flow of heat out of the volume. To summarize, the kinetic energy flux equals the sum of the kinetic energy density times velocity plus the velocity times the pressure, which actually represents the work to push the volume plus the loss of energy due to flow of heat out. In mathematical terms this is expressed as

$$\mathbf{F}_W = \mathbf{v} W + nk_B \vec{T} \cdot \mathbf{v} + \mathbf{Q} . \quad (16)$$

With the above considerations, the momentum and the energy balance equations reduce to

$$\begin{aligned}\frac{\partial J_z}{\partial t} &= \frac{2e}{m^*} \sum_i \frac{\partial}{\partial x_i} \left( K_{iz} + \frac{1}{2} n k_B T_{iz} \right) + \frac{ne^2}{m^*} E_z - \left\langle \left\langle \frac{1}{\tau_m} \right\rangle \right\rangle J_z \\ \frac{\partial W}{\partial t} &= -\nabla \cdot (\mathbf{v}W + \mathbf{Q} + n k_B \vec{T} \cdot \mathbf{v}) + \mathbf{E} \cdot \mathbf{J}_n - \left\langle \left\langle \frac{1}{\tau_E} \right\rangle \right\rangle (W - W_0)\end{aligned}\quad (17)$$

For displaced-Maxwellian approximation for the distribution function, the heat flux  $\mathbf{Q} = 0$ . However, Blotekjaer [21] has pointed out that this term must be significant for non-Maxwellian distributions, so that a phenomenological description for the heat flux, of the form described by Franz-Wiedermann law, which states that

$$\mathbf{Q} = -\kappa \nabla T_c \quad (18)$$

is used, where  $\kappa$  is the thermal or heat conductivity. In silicon, the experimental value of  $\kappa$  is 142.3 W/mK. The above description for  $\mathbf{Q}$  actually leads to a closed set of equations in which the energy balance equation is of the form

$$\begin{aligned}\frac{\partial W}{\partial t} &= -\nabla \cdot (\mathbf{v}W - \kappa \nabla T_c + n k_B T_c \mathbf{v}) + \mathbf{E} \cdot \mathbf{J}_n \\ &\quad - \left\langle \left\langle \frac{1}{\tau_E} \right\rangle \right\rangle (W - W_0)\end{aligned}\quad (19)$$

It has been recognized in recent years that this approach is not correct for semiconductors in the junction regions, where high and unphysical velocity peaks are established by the Franz-Wiedemann law. To avoid this problem, Stettler, Alam and Lundstrom [22] have suggested a new form of closure

$$\mathbf{Q} = -\kappa \nabla T_c + \frac{5}{2} (1-r) \frac{k_B T_L}{e} \mathbf{J} \quad (20)$$

where  $\mathbf{J}$  is the current density and  $r$  is a tunable parameter less than unity. Now using

$$\begin{aligned}\frac{\partial}{\partial x} (2K_{iz}) &= \frac{\partial}{\partial x_i} (n m^* v_{di} v_{dz}) = n m^* \frac{\partial}{\partial x} (v_{di} v_{dz}) \\ &= n m^* \left[ \frac{\partial v_{di}}{\partial x_i} v_{dz} + v_{dz} \frac{\partial v_{dz}}{\partial x_z} \right]\end{aligned}\quad (21)$$

and assuming that the spatial variations are confined along the z-direction, we have

$$\frac{\partial}{\partial x_z} (2K_{iz}) = \frac{\partial}{\partial x_z} (n m^* v_{dz}^2). \quad (22)$$

To summarize, the balance equations for the drifted-Maxwellian distribution function simplify to

$$\begin{aligned}
\frac{\partial n}{\partial t} &= \frac{1}{e} \nabla \cdot J_n + S_n \\
\frac{\partial J_z}{\partial t} &= \frac{e}{m^*} \frac{\partial}{\partial x_z} (nm^* v_{dz}^2 + nk_B T_c) \\
&\quad + \frac{ne^2}{m^*} E_z - \left\langle \left\langle \frac{1}{\tau_m} \right\rangle \right\rangle J_z \\
\frac{\partial W}{\partial t} &= - \frac{\partial}{\partial x_z} \left[ (W + nk_B T_c) v_{dz} - \kappa \frac{\partial T_c}{\partial x_z} \right] \\
&\quad + J_z E_z - \left\langle \left\langle \frac{1}{\tau_E} \right\rangle \right\rangle (W - W_0)
\end{aligned} \tag{23}$$

where

$$\begin{aligned}
J_z &= -env_{dz} = -\frac{e}{m^*} P_z \\
W &= \frac{1}{2} nm^* v_{dz}^2 + \frac{3}{2} nk_B T_c
\end{aligned} \tag{24}$$

### 2.3 Failure of the Drift-Diffusion and hydrodynamic models

To understand the advantages and the limitations of the drift-diffusion and of the hydrodynamic model, let us consider the following examples: Fully Depleted (FD) SOI devices with channel lengths 25, 45 and 90 nm. The oxide thickness and the doping of the channel of the three devices considered are summarized in Table 1. In Figures 4 and 5 (a-c) we compare the output characteristics of the three devices when using the drift-diffusion and the hydrodynamic model.

Feature (channel length)	14 nm	25 nm	90 nm
Tox	1 nm	1.2 nm	1.5 nm
V <sub>DS</sub>	1V	1.2 V	1.4 V
Overshoot EB/HD without series resistance	233% / 224%	139% / 126%	31% / 21%
Overshoot EB/DD with series resistance	153% / 96%	108% / 67%	39% / 26%

Source/drain doping =  $10^{20} \text{ cm}^{-3}$  and  $10^{19} \text{ cm}^{-3}$  (series resistance (SR) case)

Channel doping =  $10^{18} \text{ cm}^{-3}$

Overshoot =  $(I_{DHD} - I_{DD}) / I_{DD}$  (%);  $I_D$  is the on-state current

Table 1. Geometrical dimensions and applied biases of the fully-depleted SOI nMOSFETs simulated here.

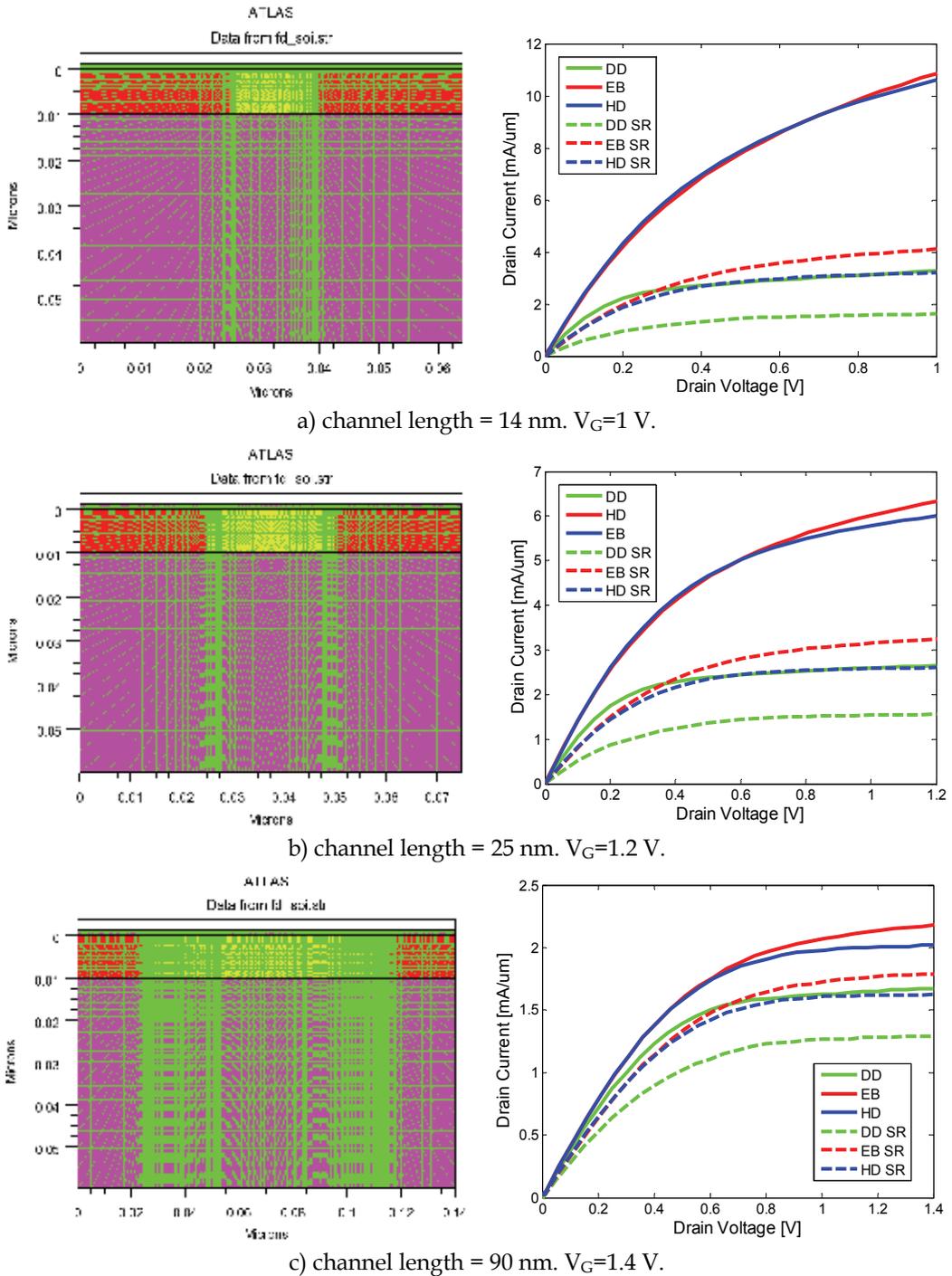
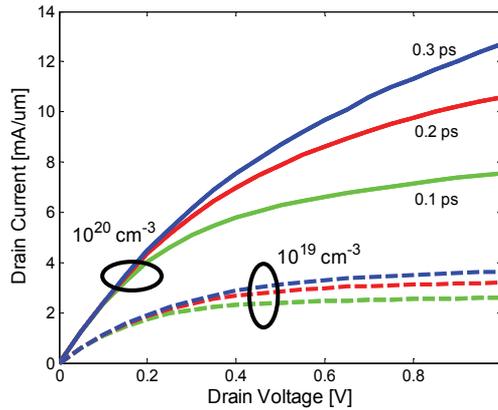
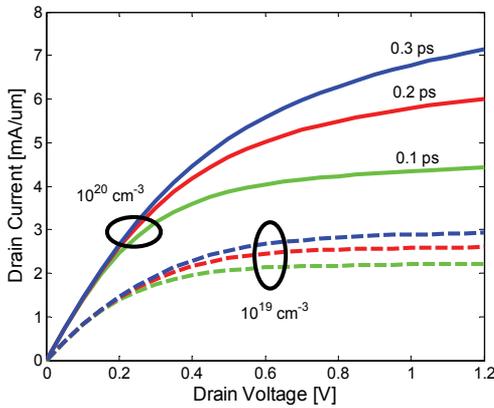


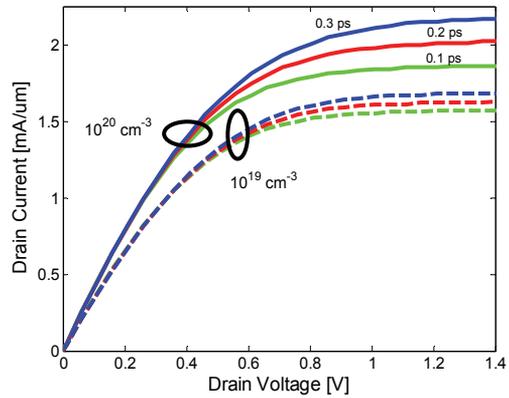
Fig. 4. Mesh and output characteristics of 14, 25 and 90 nm channel length FD SOI devices in the on-state when using drift-diffusion, energy balance, and hydrodynamic models. SR stands for series resistance.



a) Simulated characteristics for different energy relaxation times for two different source/drain doping for a channel length of 14 nm,  $V_G=1$  V.



b) Simulated characteristics for different energy relaxation times for two different source/drain doping, for a channel length of 25 nm,  $V_G=1.2$  V.



c) Simulated characteristics for different energy relaxation times for two different source/drain doping, for a channel length of 90 nm,  $V_G=1.4$  V.

Fig. 5. Dependence of the on-state current upon the choice of the energy relaxation time for three different channel length FD SOI devices.

Here we use the commercial Silvaco Atlas (PISCEC) simulation package [23] that includes hydrodynamic modeling with momentum and energy relaxation times of 0.2 ps, Auger generation/recombination (important for the proper modeling of the heavily doped source and drain contacts), and the Shockley-Read-Hall (SRH) generation-recombination mechanism are included here for completeness, although the latter is not really important for this device structure. Impact ionization is not included in these simulations. In the hydrodynamic calculation, it is important that one uses the NEWTON method for solving the coupled set of equations, otherwise the simulation will not converge due to the strong coupling of the equations at high drain biases. We consider both the simplified energy balance (EB) model and the complete hydrodynamic model (HD). We present simulation results for the following two cases:

1. Source and drain doping of  $10^{20}$  and  $10^{19}$   $\text{cm}^{-3}$  to examine series resistance effects. This is very important to know as in prototypical Monte Carlo device simulations source and drain regions are usually doped up to  $10^{19}$   $\text{cm}^{-3}$  to reduce the computational cost (total number of particles simulated). In these simulations we assume that the energy relaxation time is 0.2 ps, which is a typical value used for the silicon material system. The results from these simulations are presented in Figure 4 for the 14 nm, 25 nm and 90 nm channel length devices. On the left panel, we show the meshing used in these simulations and on the right panel we show the output characteristics for the appropriate on-state gate bias and drift-diffusion and hydrodynamic transport models.
2. In this second case we perform only hydrodynamic simulations to investigate the sensitivity of the hydrodynamic model to variations in the energy relaxation time which, in principle, is a material and device geometry dependent parameter which makes it almost impossible to determine analytically. This variation for the three technology nodes of devices is shown in Figure 5.

The results first show that the source/drain doping plays an important role in terms of the drive current, which is primarily effect of series resistance. From the results presented it is evident that non-stationary transport plays smaller role in 90 nm gate-length FD SOI devices, whereas the importance of non-stationary transport and the velocity overshoot associated with it increases drastically for 14 nm gate length FD SOI device. These results suggest that one must include energy balance equation if proper modeling of nano-scale devices with gate lengths less than 100 nm is to be achieved.

Yet another issue that deserves further attention is the dependence of the simulation results upon the choice of the energy relaxation time. In Figure 5 we plot the output characteristics of 14, 25 and 90 nm gate length FD SOI devices in which parameter is the energy relaxation time. We see strong dependence of the on-current upon the choice of the energy relaxation time for the smallest structure being simulated which suggests that proper determination of the energy relaxation time is needed. The energy relaxation time, in turn, is bias and geometry dependent parameter and its exact determination is impossible. The inability to properly determine the energy relaxation time in hydrodynamic/energy balance models has been the main motivation for the development of particle-based simulators discussed in Chapters 3 and 4.

### 3. Bulk Monte Carlo method

In the previous section we have considered continuum methods of describing transport in semiconductors, specifically the drift-diffusion and hydrodynamic models, which are derived from moments of the semi-classical Boltzmann Transport Equation (BTE). As approximations to the BTE, it was shown that in the case of small devices (see Section 2.3 above), such approaches become inaccurate, or fail completely. Indeed, one can envision that, as physical dimensions are reduced, at some level a continuum description of current breaks down, and the granular nature of the individual charge particles constituting the charge density in the active device region becomes important.

The microscopic simulation of the motion of individual particles in the presence of the forces acting on them due to external fields as well as the internal fields of the crystal lattice and other charges in the system has long been popular in the chemistry community, where *molecular dynamics* simulation of atoms and molecules have long been used to investigate the thermodynamic properties of liquids and gases. In solids, such as semiconductors and

metals, transport is known to be dominated by random scattering events due to impurities, lattice vibrations, etc., which randomize the momentum and energy of charge particles in time. Hence, stochastic techniques to model these random scattering events are particularly useful in describing transport in semiconductors, in particular the *Monte Carlo* method.

The Ensemble Monte Carlo techniques have been used for well over 30 years as a numerical method to simulate nonequilibrium transport in semiconductor materials and devices and has been the subject of numerous books and reviews [24,25,26]. In application to transport problems, a random walk is generated using the random number generating algorithms common to modern computers, to simulate the stochastic motion of particles subject to collision processes. This process of random walk generation is part of a very general technique used to evaluate integral equations and is connected to the general random sampling technique used in the evaluation of multi-dimensional integrals [27].

The basic technique as applied to transport problems is to simulate the free particle motion (referred to as the *free-flight*) terminated by instantaneous random *scattering events*. The Monte Carlo algorithm consists of generating random free flight times for each particle, choosing the type of scattering occurring at the end of the free flight, changing the final energy and momentum of the particle after scattering, and then repeating the procedure for the next free flight. Sampling the particle motion at various times throughout the simulation allows for the statistical estimation of physically interesting quantities such as the single particle distribution function, the average drift velocity in the presence of an applied electric field, the average energy of the particles, etc. By simulating an *ensemble* of particles, representative of the physical system of interest, the non-stationary time-dependent evolution of the electron and hole distributions under the influence of a time-dependent driving force may be simulated.

This particle-based picture, in which the particle motion is decomposed into free flights terminated by instantaneous collisions, is basically the same approximate picture underlying the derivation of the semi-classical Boltzmann Transport Equation (BTE). In fact, it may be shown that the one-particle distribution function obtained from the random walk Monte Carlo technique satisfies the BTE for a homogeneous system in the long-time limit [28]. This semi-classical picture breaks down when quantum mechanical effects become pronounced, and one cannot unambiguously describe the instantaneous position and momentum of a particle. In the following, we first describe the derivation of the free-flight scatter sequence using the path-integral method (section 3.1) and then we describe the standard Monte Carlo algorithm used to simulate charge transport in semiconductors (section 3.2). We then discuss how this basic model for charge transport within the BTE is self-consistently solved with the appropriate field equations to perform particle based device simulation (section 4).

### 3.1 Monte Carlo and path-integral methods

The path-integral method for solving the BTE is a rather a useful and an intuitive procedure for describing the Monte Carlo method. In its general form the BTE is:

$$\begin{aligned} \frac{\partial f}{\partial t} + \vec{v} \cdot \nabla_r f + (-e)\vec{E} \cdot \nabla_p f &= \left. \frac{\partial f}{\partial t} \right|_{coll} = \\ &= \sum_{i=1}^N \sum_{\vec{p}'} [S_i(\vec{p}', \vec{p}) f(\vec{r}, \vec{p}', t) - S_i(\vec{p}, \vec{p}') f(\vec{r}, \vec{p}, t)] \end{aligned} \quad (25)$$

The first term on the right-hand side (RHS) of Eq. (25) gives the scattering into state  $\vec{p}$ , while the second term on the RHS of Eq. (25) is the scattering out of state  $\vec{p}$ . This form of the BTE is valid for non-degenerate semiconductors. The collision integral on the RHS can also be expressed as:

$$RHS = \sum_{\vec{p}'} \sum_{i=1}^N [S_i(\vec{p}', \vec{p}) f(\vec{r}, \vec{p}', t)] - f(\vec{r}, \vec{p}, t) \sum_{\vec{p}'} \sum_{i=1}^N S_i(\vec{p}, \vec{p}'), \quad (26)$$

where,  $\frac{1}{\tau(\vec{p})} = \sum S_i(\vec{p}, \vec{p}')$  is the total scattering rate out of state  $\vec{p}$ .

Hence:

$$\begin{aligned} RHS &= \sum_{\vec{p}'} \sum_{i=1}^N [S_i(\vec{p}', \vec{p}) f(\vec{r}, \vec{p}', t)] - f(\vec{r}, \vec{p}, t) \left[ \frac{1}{\tau(\vec{p})} + \Omega(\vec{p}) \right] + f(\vec{r}, \vec{p}, t) \Omega(\vec{p}) = \\ &= \sum_{\vec{p}'} \sum_{i=1}^N [S_i(\vec{p}', \vec{p}) f(\vec{r}, \vec{p}', t)] - \Gamma(\vec{p}) f(\vec{r}, \vec{p}, t) + \sum_{\vec{p}'} f(\vec{r}, \vec{p}', t) \Omega(\vec{p}) \delta_{\vec{p}, \vec{p}'} \end{aligned} \quad (27)$$

In this last expression we have added a term  $\Omega(\vec{p})$  such that the total scattering rate out of a state  $p$  is constant. To understand the meaning of this term we need to go backwards, i.e. write  $\Gamma(\vec{p})$  as:

$$\begin{aligned} \Gamma(\vec{p}) &= \frac{1}{\tau(\vec{p})} + \Omega(\vec{p}) = \sum_{\vec{p}'} \sum_{i=1}^N S_i(\vec{p}, \vec{p}') + \sum_{\vec{p}'} \Omega(\vec{p}) \delta_{\vec{p}, \vec{p}'} = \\ &= \sum_{\vec{p}'} \left[ \sum_{i=1}^N S_i(\vec{p}, \vec{p}') + \Omega(\vec{p}) \delta_{\vec{p}, \vec{p}'} \right] \end{aligned} \quad (28)$$

We now define an effective transition rate:

$$S_{eff}(\vec{p}, \vec{p}') = \sum_{i=1}^N S_i(\vec{p}, \vec{p}') + \Omega(\vec{p}) \delta_{\vec{p}, \vec{p}'}, \quad (29)$$

which consists of the sum of the  $N$  physical transition rates plus a term that has a momentum conserving  $\delta$ -function. This second term has no effect on the carrier momentum and energy and it is a fictitious scattering process which is called self-scattering. The self-scattering can be calculated from:

$$\Omega(\vec{p}) = \Gamma(\vec{p}) - \frac{1}{\tau(\vec{p})}. \quad (30)$$

With the above definition for the self-scattering term, the BTE becomes:

$$\frac{\partial f}{\partial t} + \vec{v} \nabla_{\vec{r}} f + (-e) \vec{E} \nabla_{\vec{p}} f + \Gamma(\vec{p}) f(\vec{r}, \vec{p}, t) = \sum_{\vec{p}'} \left[ \sum_{i=1}^N S_i(\vec{p}', \vec{p}) + \left( \Gamma(\vec{p}) - \frac{1}{\tau(\vec{p})} \right) \delta_{\vec{p}, \vec{p}'} \right] f(\vec{r}, \vec{p}', t). \quad (31)$$

For homogenous samples, the BTE reduces to:

$$\begin{aligned} \frac{\partial f}{\partial t} + \bar{v} \nabla_r f + (-e) \bar{\varepsilon} \nabla_p f + \Gamma(\bar{p}) f(\bar{r}, \bar{p}, t) = \tilde{I}(\bar{p}, t) = \\ = \sum_{\bar{p}'} \left[ \sum_{i=1}^N S_i(\bar{p}', \bar{p}) + \left( \Gamma(\bar{p}) - \frac{1}{\tau(\bar{p})} \right) \delta_{\bar{p}, \bar{p}'} \right] f(\bar{p}', t) \end{aligned} \quad (32)$$

In the last formulation of the BTE, the coordinate space (phase space) is fixed and the electrons move along given trajectory in response to the applied forces. With the introduction of variables:

$$\begin{cases} \tilde{t} = t \\ \tilde{p} = \bar{p} + e \bar{\varepsilon} t' \end{cases} \quad (33)$$

we go to a description in which electrons are frozen in their positions and the coordinate system is moving. Then,

$$\frac{\partial f}{\partial t} = \frac{\partial f}{\partial \tilde{t}} \cdot \frac{\partial \tilde{t}}{\partial t} + \frac{\partial f}{\partial \tilde{p}} \cdot \frac{\partial \tilde{p}}{\partial t} = \frac{\partial f}{\partial \tilde{t}} + e \bar{\varepsilon} \cdot \frac{\partial f}{\partial \tilde{p}}, \quad (34a)$$

$$\frac{\partial f}{\partial \tilde{p}} = \frac{\partial f}{\partial \tilde{t}} \cdot \frac{\partial \tilde{t}}{\partial \tilde{p}} + \frac{\partial f}{\partial \tilde{p}} \frac{\partial \tilde{p}}{\partial \tilde{p}} = \frac{\partial f}{\partial \tilde{p}}. \quad (34b)$$

In this notation, the BTE becomes:

$$\frac{\partial f}{\partial \tilde{t}} + \Gamma \cdot f(\tilde{p} - e \bar{\varepsilon} \tilde{t}, \tilde{t}) = \tilde{I}(\tilde{p} - e \bar{\varepsilon} \tilde{t}, \tilde{t}). \quad (35)$$

The solution of the homogenous equation of the form:

$$\frac{\partial f}{\partial \tilde{t}} + \Gamma \cdot f(\tilde{p} - e \bar{\varepsilon} \tilde{t}, \tilde{t}) = 0, \quad (36)$$

can be found using a separation of variables method, to be:

$$\ln f(\tilde{p} - e \bar{\varepsilon} \tilde{t}, \tilde{t}) \Big|_0^{\tilde{t}} = -\Gamma \int_0^{\tilde{t}} d\tilde{t}, \quad (37a)$$

or:

$$\ln \left[ f(\tilde{p} - e \bar{\varepsilon} \tilde{t}, \tilde{t}) / f(\tilde{p}, 0) \right] = -\Gamma \tilde{t}, \quad (37b)$$

to get:

$$f(\tilde{p} - e \bar{\varepsilon} \tilde{t}, \tilde{t}) = f(\tilde{p}, 0) e^{-\Gamma \tilde{t}}. \quad (38)$$

Going back to the original coordinate system gives:

$$f(\bar{p}, t) = f(\bar{p} + e \bar{\varepsilon} t, 0) e^{-\Gamma t}. \quad (39)$$

This term is the transient term. It states that an electron initially in a state  $(\vec{p} + e\vec{\varepsilon}t)$  at time  $t=0$ , has arrived in a state  $\vec{p}$  at time  $t$  without scattering. This event occurs with a transition probability:

$$P(\vec{p}, t, 0) = e^{-\Gamma t}, \text{ for } \Gamma(\vec{p}) = \Gamma = \text{const.} \quad (40)$$

For general case, when  $\Gamma(\vec{p})$  is not a constant, one would have had:

$$P(\vec{p}, t, 0) = \exp \left[ -\int_0^t \Gamma(\vec{p}) dt' \right]. \quad (41)$$

Also, if the initial momentum of the electron is  $(\vec{p} + e\vec{\varepsilon}t)$ , because all of the drift motion and the acceleration by the electric field, the final electron momentum at time  $t$  equals to:  $\vec{p}' = \vec{p} + e\vec{\varepsilon}t - e\vec{\varepsilon}t = \vec{p}$ .

The next task is to find a solution of the BTE for homogenous systems. The homogenous solution suggests that the general solution will also involve an exponentials. For this purpose we define a function:

$$f_1(\vec{p} - e\vec{\varepsilon}\tilde{t}, \tilde{t}) = f(\vec{p} - e\vec{\varepsilon}\tilde{t}, \tilde{t})e^{\Gamma\tilde{t}}, \quad (42a)$$

which leads to:

$$f(\vec{p} - e\vec{\varepsilon}\tilde{t}, \tilde{t}) = f_1(\vec{p} - e\vec{\varepsilon}\tilde{t}, \tilde{t})e^{-\Gamma\tilde{t}}. \quad (42b)$$

Then:

$$\frac{\partial f}{\partial \tilde{t}} = \frac{\partial f_1}{\partial \tilde{t}} e^{-\Gamma\tilde{t}} + \Gamma \cdot f_1(\vec{p} - e\vec{\varepsilon}\tilde{t}, \tilde{t})e^{-\Gamma\tilde{t}}. \quad (43)$$

Substituting this result back into the BTE gives:

$$\frac{\partial f_1}{\partial \tilde{t}} = e^{\Gamma\tilde{t}} \tilde{I}(\vec{p} - e\vec{\varepsilon}\tilde{t}, \tilde{t}). \quad (44)$$

Solving the last equation for  $f_1$  finally leads to:

$$f_1(\vec{p} - e\vec{\varepsilon}\tilde{t}, \tilde{t}) \Big|_0^{\tilde{t}} = \int_0^{\tilde{t}} d\tilde{t}_1 \cdot e^{\Gamma\tilde{t}_1} \tilde{I}(\vec{p} - e\vec{\varepsilon}\tilde{t}_1, \tilde{t}_1), \quad (45a)$$

or:

$$f_1(\vec{p} - e\vec{\varepsilon}\tilde{t}, \tilde{t}) = f_1(\vec{p}, 0) + \int_0^{\tilde{t}} d\tilde{t}_1 \cdot e^{\Gamma\tilde{t}_1} \tilde{I}(\vec{p} - e\vec{\varepsilon}\tilde{t}_1, \tilde{t}_1). \quad (45b)$$

Multiplying the last equation by  $e^{-\Gamma\tilde{t}}$  one gets:

$$f(\vec{p} - e\vec{\varepsilon}\tilde{t}, \tilde{t}) = f(\vec{p}, 0) \cdot e^{-\Gamma\tilde{t}} + \int_0^{\tilde{t}} d\tilde{t}_1 \cdot e^{-\Gamma(\tilde{t}-\tilde{t}_1)} \tilde{I}(\vec{p} - e\vec{\varepsilon}\tilde{t}_1, \tilde{t}_1). \quad (46)$$

Returning to the original coordinate system gives:

$$f(\bar{p}, t) = f(\bar{p} + e\bar{\varepsilon}t, 0)e^{-\Gamma t} + \int_0^t dt_1 \cdot e^{-\Gamma(t-t_1)} \tilde{I}(\bar{p}, t_1), \quad (47)$$

where  $\tilde{I}(\bar{p}, t) = \sum_{\bar{p}'} S_{eff}(\bar{p}', \bar{p}) f(\bar{p}', t)$ . Substituting this back gives:

$$f(\bar{p}, t) = f(\bar{p} + e\bar{\varepsilon}t, 0)e^{-\Gamma t} + \int_0^t dt_1 \sum_{\bar{p}'} f(\bar{p}', t_1) S_{eff}(\bar{p}', \bar{p} + e\bar{\varepsilon}(t-t_1)) e^{-\Gamma(t-t_1)}, \quad (48)$$

where:

- $f(\bar{p} + e\bar{\varepsilon}t, 0)e^{-\Gamma t}$  is the transient term;
- $f(\bar{p}', t_1)$  is the probability that at time  $t_1$  a state  $\bar{p}'$  is occupied by an electron;
- $S_{eff}(\bar{p}', \bar{p} + e\bar{\varepsilon}(t-t_1))$  is the transition rate (probability) from state  $\bar{p}'$  to state  $\bar{p} + e\bar{\varepsilon}(t-t_1)$ .
- $e^{-\Gamma(t-t_1)}$  is the probability that an electron will not undergo collision event in interval  $(t-t_1)$ .

This last expression is known as Chambers-Rees path integral. Rees [29] innovation is the introduction of the fictitious scattering term. Ignoring the transient term, one can find the solution of the distribution function using the following iterative procedure that is obtained by time discretization, i.e. using  $t=N \cdot \Delta t$  and  $t_n=n \cdot \Delta t$ . Then,

$$f_N(\bar{p}) = \Delta t \sum_{m=0}^{N-1} \sum_{\bar{p}'} f_m(\bar{p}') S_{eff}(\bar{p}', \bar{p} + e\bar{\varepsilon}(N-m)\Delta t) e^{-\Gamma(N-m)\Delta t}. \quad (49)$$

The two step procedure is then found by using  $N=1$ , which means that  $t=\Delta t$ , i.e.,

$$f_1(\bar{p}) = \Delta t \sum_{\bar{p}'} f_0(\bar{p}') S_{eff}(\bar{p}', \bar{p} + e\bar{\varepsilon}\Delta t) e^{-\Gamma\Delta t}, \quad (50)$$

where,

- $g_0(\bar{p} + e\bar{\varepsilon}\Delta t) = f_0(\bar{p}') \cdot S_{eff}(\bar{p}', \bar{p} + e\bar{\varepsilon}\Delta t)$  is the intermediate function that describes the occupancy of a state  $\bar{p} + e\bar{\varepsilon}\Delta t$  at time  $t=0$ , which can be changed due to in-scattering events;
- $e^{-\Gamma\Delta t}$  is the probability that no scattering occurred within time integral  $\Delta t$  (free-flight).

Now assume that  $t=2\Delta t$ . This then gives:

$$\begin{aligned} f_2(\bar{p}) &= \Delta t \sum_{m=0}^1 \sum_{\bar{p}'} f_m(\bar{p}') S_{eff}(\bar{p}', \bar{p} + e\bar{\varepsilon}(2-m)\Delta t) e^{-\Gamma(2-m)\Delta t} = \\ &= \Delta t \left\{ \sum_{\bar{p}'} f_0(\bar{p}') S_{eff}(\bar{p}', \bar{p} + e\bar{\varepsilon}(2\Delta t)) e^{-\Gamma(2\Delta t)} + \right. \\ &\quad \left. + \sum_{\bar{p}'} f_1(\bar{p}') S_{eff}(\bar{p}', \bar{p} + e\bar{\varepsilon}\Delta t) e^{-\Gamma\Delta t} \right\} \end{aligned} \quad (51)$$

These examples suggest that the evaluation of  $f_{n+1}(\vec{p})$  involves integration over trajectories and the exponential factors just give the probability that no scattering has occurred.

### 3.2 Bulk Monte Carlo method

According to the description provided in Section 3.1 above, in the bulk Monte Carlo method, particle motion is assumed to consist of free flights terminated by instantaneous scattering events, which change the momentum and energy of the particle after scattering. So, the first task is to generate free flights of random time duration for each particle. To simulate this process, the probability density,  $P(t)$ , is required, in which  $P(t)dt$  is the joint probability that a particle will arrive at time  $t$  without scattering after a previous collision occurring at time  $t = 0$ , and then suffer a collision in a time interval  $dt$  around time  $t$ . The probability of scattering in the time interval  $dt$  around  $t$  may be written as  $\Gamma[\mathbf{k}(t)]dt$ , where  $\Gamma[\mathbf{k}(t)]$  is the scattering rate of an electron or hole of wavevector  $\mathbf{k}$ . The scattering rate,  $\Gamma[\mathbf{k}(t)]$ , represents the sum of the contributions from each individual scattering mechanism, which are usually calculated quantum mechanically using perturbation theory, as described later. The implicit dependence of  $\Gamma[\mathbf{k}(t)]$  on time reflects the change in  $\mathbf{k}$  due to acceleration by internal and external fields. For electrons subject to time independent electric and magnetic fields, the time evolution of  $\mathbf{k}$  between collisions is represented as

$$\mathbf{k}(t) = \mathbf{k}(0) - \frac{e(\mathbf{E} + \mathbf{v} \times \mathbf{B})t}{\hbar}, \quad (52)$$

where  $\mathbf{E}$  is the electric field,  $\mathbf{v}$  is the electron velocity and  $\mathbf{B}$  is the magnetic flux density. In terms of the scattering rate,  $\Gamma[\mathbf{k}(t)]$ , the probability that a particle has not suffered a collision after a time  $t$  is given by  $\exp\left[-\int_0^t \Gamma[\mathbf{k}(t')]dt'\right]$ . Thus, the probability of scattering in the time interval  $dt$  after a free flight of time  $t$  may be written as the joint probability

$$P(t)dt = \Gamma[\mathbf{k}(t)]\exp\left[-\int_0^t \Gamma[\mathbf{k}(t')]dt'\right]dt. \quad (53)$$

Random flight times may be generated according to the probability density  $P(t)$  above using, for example, the pseudo-random number generator implicit on most modern computers, which generate uniformly distributed random numbers in the range [0,1]. Using a direct method (see, for example [24]), random flight times sampled from  $P(t)$  may be generated according to

$$r = \int_0^{t_r} P(t)dt, \quad (54)$$

where  $r$  is a uniformly distributed random number and  $t_r$  is the desired free flight time. Integrating Eq. (54) with  $P(t)$  given by Eq. (53) above yields

$$r = 1 - \exp\left[-\int_0^{t_r} \Gamma[\mathbf{k}(t')]dt'\right]. \quad (55)$$

Since  $1-r$  is statistically the same as  $r$ , Eq. (55) may be simplified to

$$-\ln r = \int_0^{t_r} \Gamma[\mathbf{k}(t')] dt'. \quad (56)$$

Eq. (56) is the fundamental equation used to generate the random free flight time after each scattering event, resulting in a random walk process related to the underlying particle distribution function. If there is no external driving field leading to a change of  $\mathbf{k}$  between scattering events (for example in ultrafast photo-excitation experiments with no applied bias), the time dependence vanishes, and the integral is trivially evaluated. As noted in the previous section, in the general case where this simplification is not possible, it is expedient to introduce the so called self-scattering method [29], in which one introduces fictitious scattering mechanism whose scattering rate always adjusts itself in such a way that the total (self-scattering plus real scattering) rate is a constant in time

$$\Gamma = \Gamma[\mathbf{k}(t')] + \Gamma_{self}[\mathbf{k}(t')], \quad (57)$$

where  $\Gamma_{self}[\mathbf{k}(t')]$  is the self-scattering rate (see the discussion in Section 3.1 above). The self-scattering mechanism itself is defined such that the final state before and after scattering is identical. Hence, it has no effect on the free flight trajectory of a particle when selected as the terminating scattering mechanism, yet results in the simplification of Eq. (56) such that the free flight is given by

$$t_r = -\frac{1}{\Gamma} \ln r. \quad (58)$$

The constant total rate (including self-scattering)  $\Gamma$ , must be chosen at the start of the simulation interval (there may be multiple such intervals throughout an entire simulation) so that it is larger than the maximum scattering encountered during the same time interval. In the simplest case, a single value is chosen at the beginning of the entire simulation (constant gamma method), checking to ensure that the real rate never exceeds this value during the simulation. Other schemes may be chosen that are more computationally efficient, and which modify the choice of  $\Gamma$  at fixed time increments [30].

The algorithm described above determines the random free flight times during which the particle dynamics is treated semi-classically. For the scattering process itself, we need the type of scattering (i.e. impurity, acoustic phonon, photon emission, etc.) which terminates the free flight, and the final energy and momentum of the particle(s) after scattering. The type of scattering which terminates the free flight is chosen using a uniform random number between 0 and  $\Gamma$ , and using this pointer to select among the relative total scattering rates of all processes including self-scattering at the final energy and momentum of the particle

$$\Gamma = \Gamma_{self}[n, \mathbf{k}] + \Gamma_1[n, \mathbf{k}] + \Gamma_2[n, \mathbf{k}] + \dots \Gamma_N[n, \mathbf{k}], \quad (59)$$

with  $n$  the band index of the particle (or subband in the case of reduced-dimensionality systems), and  $\mathbf{k}$  the wavevector at the end of the free-flight. This process is illustrated schematically in Figure 6.

Once the type of scattering terminating the free flight is selected, the final energy and momentum (as well as band or subband) of the particle due to this type of scattering must

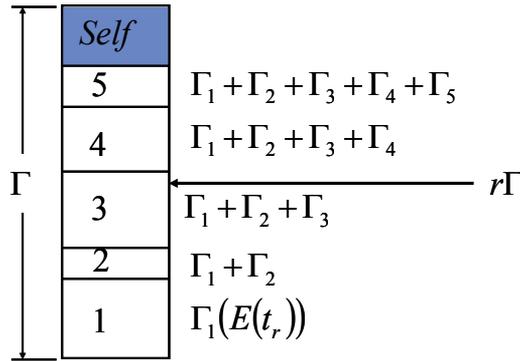


Fig. 6. Selection of the type of scattering terminating a free flight in the Monte Carlo algorithm.

be selected. For elastic scattering processes such as ionized impurity scattering, the energy before and after scattering is the same. For the interaction between electrons and the vibrational modes of the lattice described as quasi-particles known as phonons, electrons exchange finite amounts of energy with the lattice in terms of emission and absorption of phonons. For determining the final momentum after scattering, the scattering rate,  $\Gamma_j[n, \mathbf{k}; m, \mathbf{k}']$  of the  $j$ th scattering mechanism is needed, where  $n$  and  $m$  are the initial and final band indices, and  $\mathbf{k}$  and  $\mathbf{k}'$  are the particle wavevectors before and after scattering. Defining a spherical coordinate system around the initial wavevector  $\mathbf{k}$ , the final wavevector  $\mathbf{k}'$  is specified by  $|\mathbf{k}'|$  (which depends on conservation of energy) as well as the azimuthal and polar angles,  $\varphi$  and  $\theta$  around  $\mathbf{k}$ . Typically, the scattering rate,  $\Gamma_j[n, \mathbf{k}; m, \mathbf{k}']$ , only depends on the angle  $\theta$  between  $\mathbf{k}$  and  $\mathbf{k}'$ . Therefore,  $\varphi$  may be chosen using a uniform random number

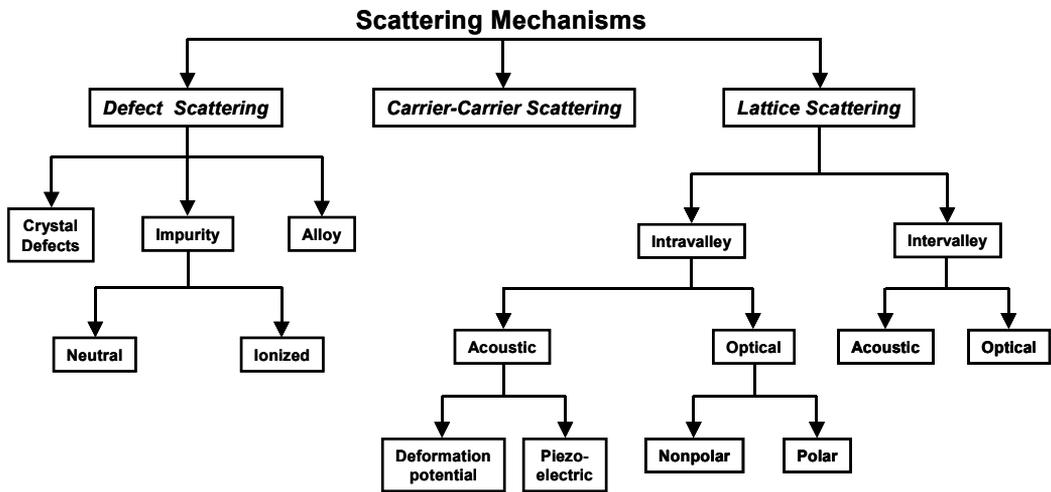


Fig. 7. Scattering mechanisms in a typical semiconductor.

between 0 and  $2\pi$  (i.e.  $2\pi r$ ), while  $\theta$  is chosen according to the angular dependence for scattering arising from  $\Gamma_j[n, \mathbf{k}; m, \mathbf{k}']$ . If the probability for scattering into a certain angle  $P(\theta)d\theta$  is integrable, then random angles satisfying this probability density may be generated from a uniform distribution between 0 and 1 through inversion of Eq. (54).

Otherwise, a rejection technique (see, for example, [24,25]) may be used to select random angles according to  $P(\theta)$ . Scattering mechanisms that contribute to transport are summarized in Figure 7. The corresponding scattering rates for general non-parabolic bands are summarized in Table 2.

A general Monte Carlo code is developed as follows. First a subroutine is typically called that contains all material and scattering rates parameters for the scattering mechanisms included in the theoretical model. After the material and run parameters are read in, in the first step of the Monte Carlo simulation procedure it is necessary to construct scattering tables for the  $\Gamma$ , L and X valleys (for GaAs as a prototypical example) that initializes a series of events that are summarized in Figure 8. At each energy, the cumulative scattering rates for each valley are stored in separate look-up tables, and renormalized according to the maximum scattering rate (including self-scattering) that occurs over the range of energies stored. The structure of these subroutines is such that adding additional scattering event has to be trivial.

1. Acoustic Phonon Scattering
$W(E) = \left( \frac{2\pi D_{ac}^2 K_B T_L}{\hbar C_l} \right) * \left( \frac{(2m_d)^{\frac{3}{2}} \sqrt{E(1+\alpha E)}}{4\pi^2 \hbar^3} \right) * (1+2\alpha E)$
2. Intervalley Phonon Scattering
$W(E) = \left( \frac{\pi D_{ij}^2 Z_j}{\rho W_{ij}} \right) * \left( n(W_{ij}) + \frac{1}{2} \mp \frac{1}{2} \right) * \left( \frac{(2m_d)^{\frac{3}{2}} \sqrt{E_f(1+\alpha E_f)}}{4\pi^2 \hbar^3} \right) * (1+2\alpha E_f)$ $E_f = E \pm \hbar W_{ij} - \Delta E_{ij}$
3. Ionized Impurity Scattering
$W(E) = \left( \frac{\sqrt{2} e^4 N_I m_d^{\frac{3}{2}}}{\pi \epsilon_s^2 \hbar^4} \right) * \left( \sqrt{E(1+\alpha E)} * (1+2\alpha E) \right) * \left( \frac{1}{q_D^2 \left( q_D^2 + \frac{8m_d E(1+\alpha E)}{\hbar^2} \right)} \right)$ $q_D = \sqrt{\frac{e^2 N_I}{\epsilon K_B T_L}}$
4. Polar Optical Phonon Scattering
$W(E) = \left( \frac{\sqrt{m_d} e^2 W_{LO}}{4\sqrt{2}\pi \hbar \epsilon_p} \right) * \left( N_o + \frac{1}{2} \mp \frac{1}{2} \right) * \left( \frac{1+2\alpha E'_k}{\gamma_k} \right) * F(E_k, E'_k)$ $N_o = \frac{1}{e^{\frac{\hbar W_{LO}}{K_B T_L}} - 1} \quad \epsilon_p = \frac{1}{\frac{1}{\epsilon_{high}} - \frac{1}{\epsilon_{low}}} \quad F(E_k, E'_k) = \ln \left( \frac{\sqrt{\gamma_k} + \sqrt{\gamma_{k'}}}{\sqrt{\gamma_k} - \sqrt{\gamma_{k'}}} \right)$ $\gamma_k = E_k(1+\alpha E_k)$ $E'_k = E_k \pm \hbar W_{LO}$

<p>5. Piezoelectric Scattering</p> $W(E) = \left( \frac{m_d^2 K_B T_L}{4\sqrt{2\pi\rho v_s^2 \hbar^2}} \right) * \left( \frac{1 + 2\alpha E}{\sqrt{E(1 + \alpha E)}} \right) * \left( \frac{ee_{pz}}{\epsilon_\infty} \right)^2 * \ln \left( 1 + \frac{8m_d E(1 + \alpha E)}{\hbar^2 q_D^2} \right)$ $q_D = \sqrt{\frac{e^2 N_l}{\epsilon K_B T_L}}$
<p>6. Dislocation Scattering (e.g. GaN)</p> $W(E) = \left( \frac{N_{dis} m_d e^4}{4\hbar^3 \epsilon^2 c^2} \right) * \left( \frac{\lambda^4}{\left( 1 + \frac{8\lambda^2 m_d E(1 + \alpha E)}{\hbar^2} \right)^{\frac{3}{2}}} \right) * \left( 1 + \frac{4\lambda^2 m_d E(1 + \alpha E)}{\hbar^2} \right) * (1 + 2\alpha E)$ $\lambda = \sqrt{\frac{\epsilon K_B T_L}{e^2 n'}}$ <p>where <math>n'</math> is the effective screening concentration  <math>N_{dis}</math> is the Line dislocation density</p>
<p>7. Alloy Disorder Scattering (<math>Al_x Ga_{1-x} As</math>)</p> $W(E) = \left( \frac{x(1-x)a^3}{\pi} \right) * \left( \frac{D_{alloy}^2 d}{\hbar^4} \right) * m_d \sqrt{2m_d E(1 + \alpha E)} * (1 + 2\alpha E)$ <p>Where: <math>d</math> is the lattice disorder (<math>0 \leq d \leq 1</math>)  <math>D_{alloy}</math> is the alloy disorder scattering potential</p>

Table 2. Scattering rates expressions for non-parabolic bands.

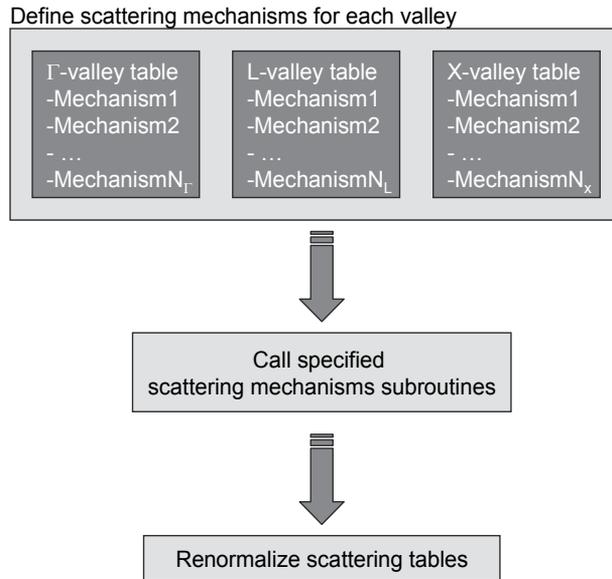


Fig. 8. Procedure for the creation of the scattering tables.

Having constructed the scattering table and after renormalizing the table, examples of which are given in Figure 9 and Figure 10 for the  $\Gamma$ , L, and X valley, the next step is to initialize carriers wavevector and energy and the initial free-flight time. This is accomplished by calling the initialization subroutine. Energy and wavevector histograms of the initial carrier energy and the components of the wave-vector along the x-, y-, and z-axes are shown in Figure 11. For good statistics, the number of particles simulated is 10000, and one can see the statistical fluctuation of these average quantities associated with the finite number of particles. Notice that the initial y-component for the wavevector is symmetric around the y-axis which means that the average wavevector along the y-axis is zero, which should be expected since the electric field along the y-component is zero at  $t=0$ . Identical distributions have been obtained for the x- and for the z-components of the wavevector. Also note that the energy distribution has the Maxwell-Boltzmann form as it should be expected. One can also estimate from this graph that the average energy of the carriers is on the order of  $(3/2)k_B T$ .

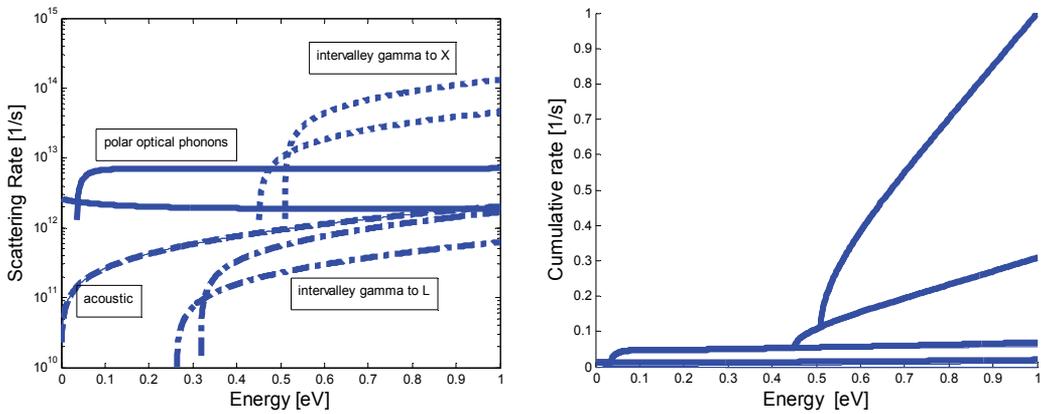


Fig. 9. Left panel: scattering rates for the  $\Gamma$ -valley. For simplicity we have omitted Coulomb scattering in these calculations. In the left figure, the dashed line corresponds to the acoustic phonon scattering rate, solid lines correspond to polar optical phonon scattering (absorption and emission), and the dashed-dotted line corresponds to intervalley scattering from  $\Gamma$ -valley to L-valley. Since the L-valley is along the [111] direction, there are 8 equivalent directions and since these valleys are shared there are a total of 4 equivalent L valleys. The dotted line corresponds to scattering from the  $\Gamma$ -valley to X-valleys. The X-valleys are at the [100] direction and since there are 6 equivalent [100] directions and the valleys are shared between Brillouin zones, there are 3 equivalent X valleys. Right panel: normalized cumulative scattering table for the  $\Gamma$ -valley. Everything above the top line up to  $\Gamma=1$  is self-scattering so it is advisable when checking the scattering mechanisms to first check whether the scattering mechanism chosen is self-scattering or not. This is in particular important for energies below 0.5 eV for this particular scattering table when the  $\Gamma$  to X intervalley scattering (absorption and emission) takes over.

When the initialization process is finished, the main free-flight-scatter procedure takes place until the completion of the simulation time. There are two components in this routine; first the carriers accelerate freely due to the electric field, accomplished by calling the **drift()** subroutine, and then their free-flights are interrupted by random scattering events that are

managed by the `scatter_carrier()` subroutine. The flow-chart for performing the free-flight-scatter process within one time step  $\Delta t$  is shown diagrammatically in Figure 12.

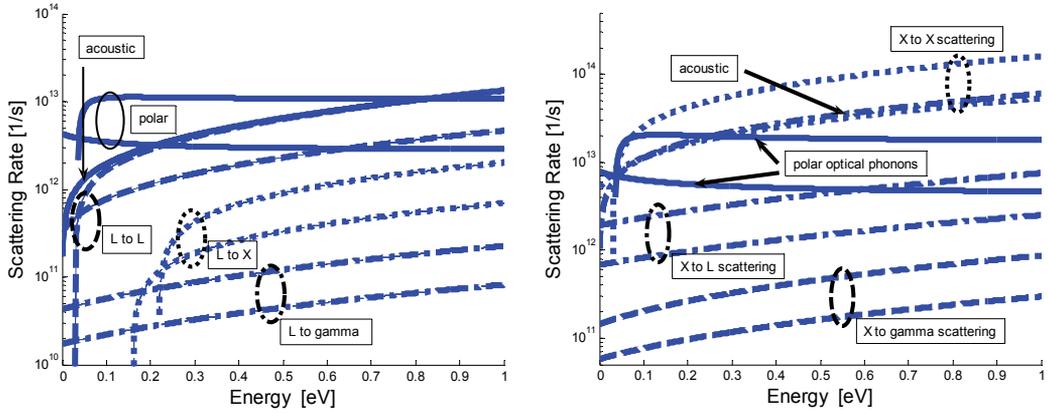


Fig. 10. Scattering rates for the L (left panel) and X (right panel) valleys used to create the corresponding normalized scattering tables (not shown here).

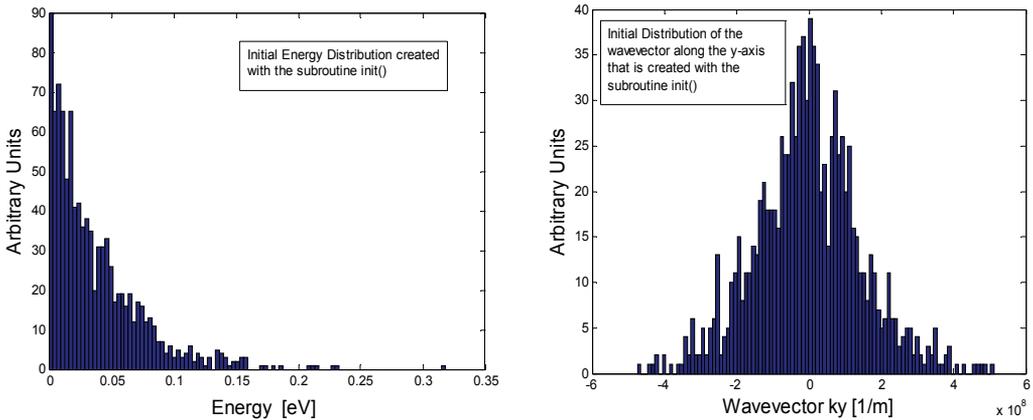


Fig. 11. Initial carrier distribution for an ensemble of 10000 Particles. Left panel: distribution of wavevector  $k_y$ . Right panel: energy distribution.

In the `scatter_carrier()` subroutine, first the scattering mechanism terminating the free flight is chosen, to which certain attributes are associated such as the change in energy after scattering. For inelastic scattering processes, we have the change in energy due to emission or absorption of phonons, for example. Also, the nature of the scattering process is identified: isotropic or anisotropic. Note that when performing acoustic phonon and intervalley scattering for GaAs, both of which are isotropic scattering processes, no coordinate system transformation is needed to determine the final wavevector after scattering. Because polar optical phonon and Coulomb scattering mechanisms are anisotropic, it is necessary to do a rotation of the coordinate system, scatter the carrier in the rotated system and then perform inverse coordinate transformation. This procedure is needed because it is much easier to determine final carrier momentum in the rotated

coordinate system in which the initial wavevector  $k$  is aligned with the  $z$ -axis. For this case, one can calculate that the final polar angle for scattering with polar optical phonons for parabolic bands in the rotated coordinate system is

$$\cos \theta = \frac{(1 + \xi) - (1 + 2\xi)^r}{\xi}, \quad \xi = \frac{2\sqrt{E_k(E_k \pm \hbar\omega_0)}}{(\sqrt{E_k} - \sqrt{E_k \pm \hbar\omega_0})^2} \quad (60)$$

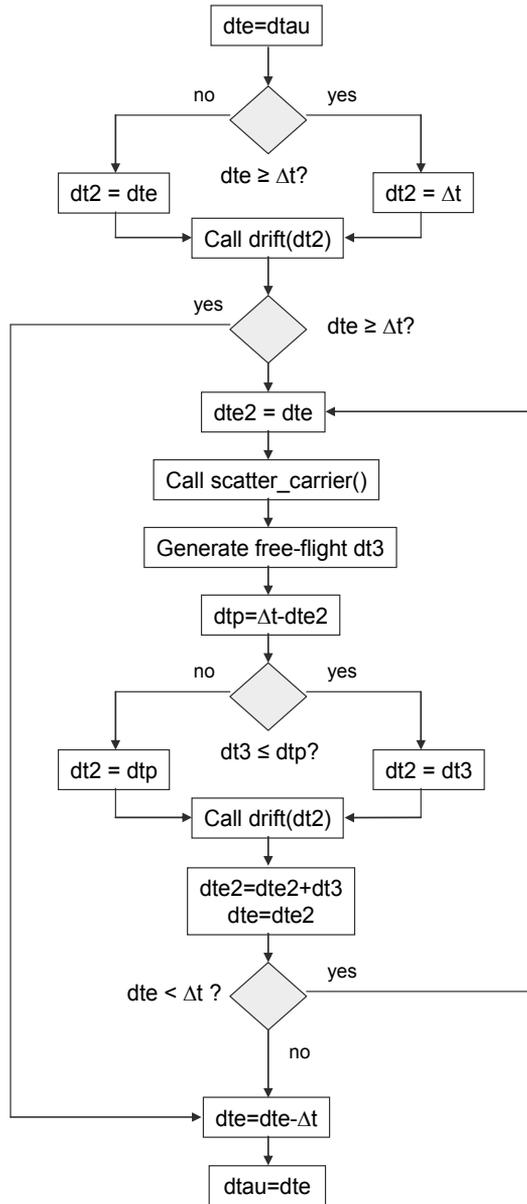


Fig. 12. Free-flight-scatter procedure within one time step.

where  $E_k$  is the carrier energy,  $\hbar\omega_0$  is the polar optical phonon energy and  $r$  is a random number uniformly distributed between 0 and 1. The final angle for scattering with ionized impurities (Coulomb scattering) and for parabolic bands is

$$\cos\theta = 1 - \frac{2r}{1 + 4k^2L_D^2(1-r)} \quad (61)$$

where  $\mathbf{k}$  is the carrier wavevector, and  $L_D$  is the Debye screening length. The azimuthal angle for both scattering processes is simply calculated using  $\varphi = 2\pi r$ . The importance of properly calculating the angle  $\theta$  after scattering to describe small angle deflections in the case of Coulomb or polar optical phonon scattering is illustrated in Figure 13 (from 0 to  $\pi=3.141592654$ ) where we plot the histogram of the polar angle after scattering for electron-polar optical phonon scattering, where we can clearly see the preference for small angle deflections that are characteristic for any Coulomb type interaction (polar optical phonon is in fact electron-dipole interaction). Graphical representation of the determination of the final angle after scattering for both isotropic and anisotropic scattering processes is given in Figure 14.

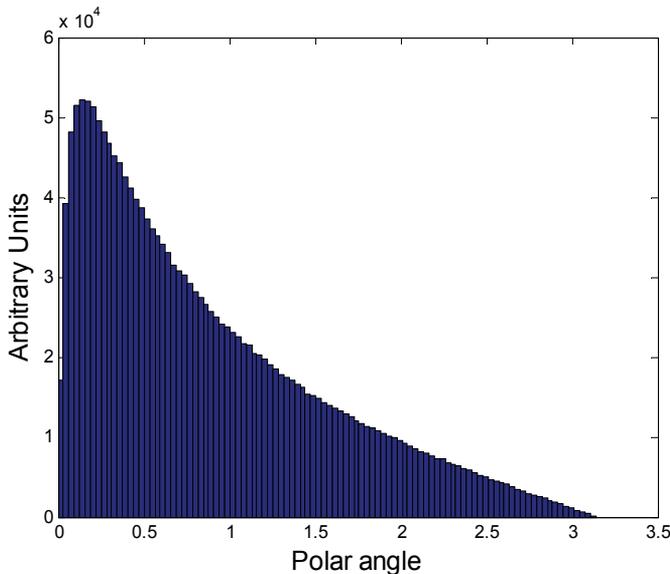


Fig. 13. Histogram of the polar angle for electron - polar optical phonon scattering.

The direct technique described above can be applied when the integrals describing  $\cos\theta$  can be analytically calculated. For most cases of interest, the integral cannot be easily inverted. In these cases a rejection technique may be employed. The procedure of the rejection technique goes as follows:

- Choose a maximum value  $C$ , such that  $C > f(x)$  for all  $x$  in the interval  $(a,b)$ .
- Choose pairs of random numbers, one between  $a$  and  $b$  ( $x_1 = a + r_1(b-a)$ ) and another  $f_1 = r_1'C$  between 0 and  $C$ , where  $r_1$  and  $r_1'$  are random numbers uniformly distributed between zero and 1.
- If  $f_1 \leq f(x_1)$ , then the number  $x_1$  is accepted as a suitable value, otherwise it is rejected.

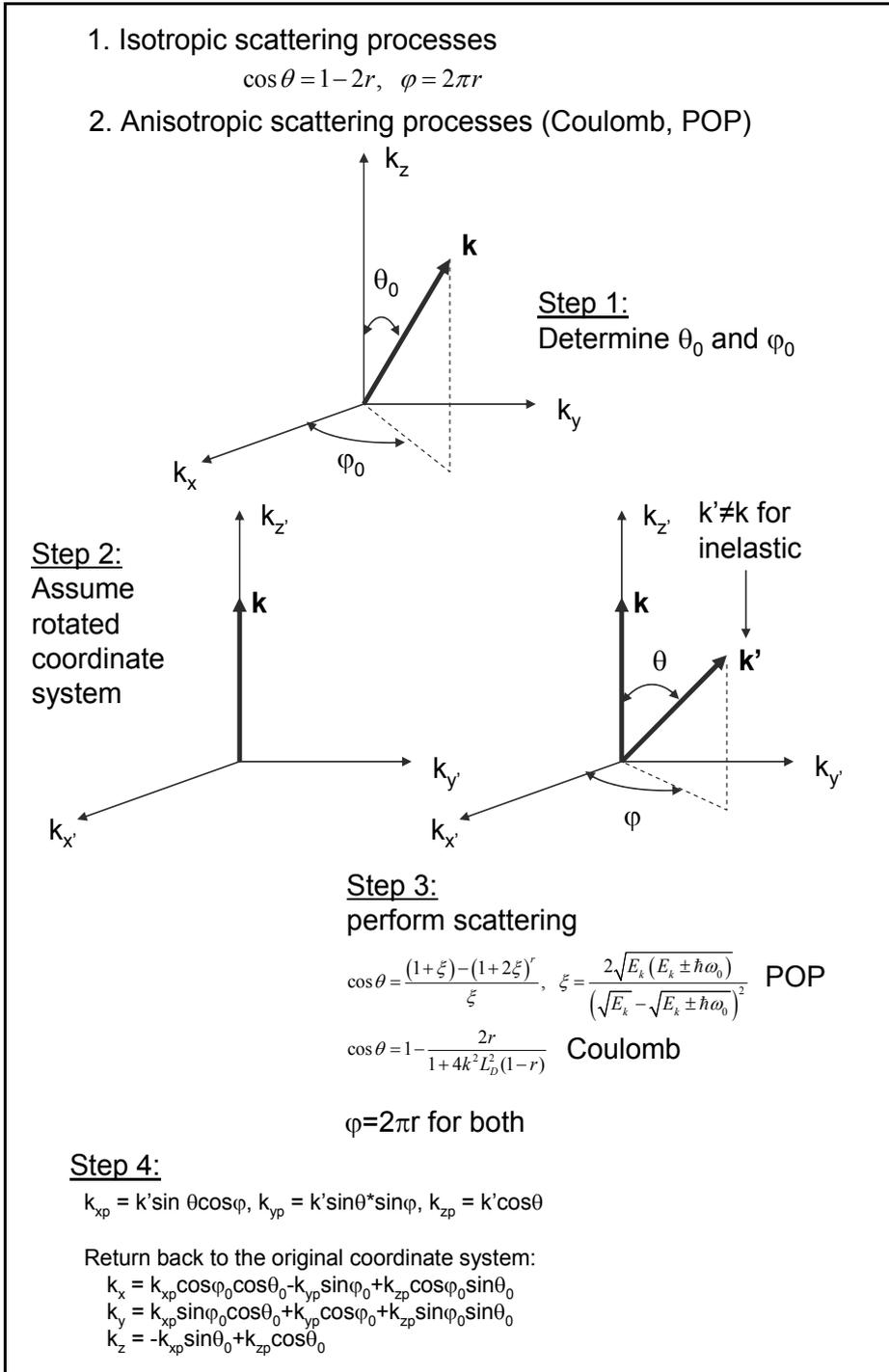


Fig. 14. Description of final angle selection for isotropic and anisotropic scattering processes using the direct technique.

The three steps described above are schematically shown in the figure below (Figure 15). For  $x = x_1$ ,  $r_1 C$  is larger than  $f(x_1)$  and in this case if this represents the final polar angle for scattering, this angle is rejected and a new sequence of two random numbers is generated to determine  $x_2$  and  $r_2 C$ . In this second case,  $f(x_2) > r_2 C$  and the polar angle  $\theta = x_2$  is selected (for polar angle selection  $a = 0$  and  $b = \pi$ ).

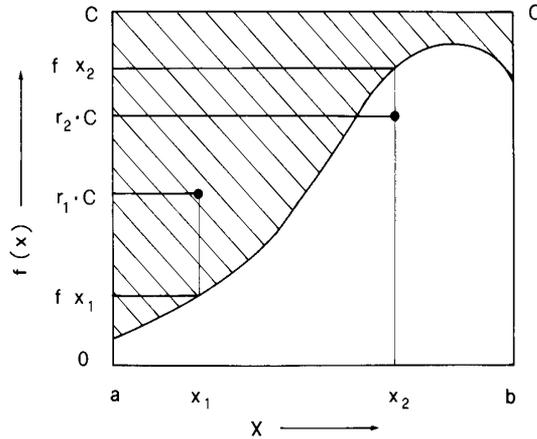


Fig. 15. Schematic description of the rejection technique.

After the simulation is completed, typical results to check are the velocity-time, the energy-time and the valley occupation versus time characteristics, such as those shown in Figure 16, where the velocity time characteristics for applied electric fields ranging from 0.5 to 7 kV/cm, with an electric field increment of 0.5 kV/cm, are shown. These clearly demonstrate that after a transient phase, the system reaches a stationary steady state, after which time we can start taking averages for calculating steady-state quantities.

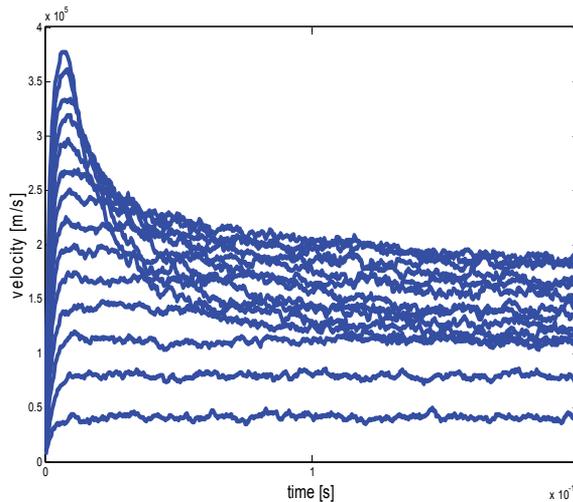


Fig. 16. Time evolution of the drift velocity for electric field strengths ranging between 0.5 and 7 kV/cm, in 0.5 kV/cm increments.

From the results shown in Figure 16, one can see that steady-state is achieved for larger time intervals when the electric field value is increased and the carriers are still sitting in the  $\Gamma$ -valley. Afterwards the time needed to get to steady-state decreases. This trend is related to the valley repopulation and movement of the carriers from the  $\Gamma$ , into the X and finally into the L valley. The steady-state velocity-field and valley population versus electric field characteristics are shown in Figure 17 and Figure 18, respectively. One can clearly see on the velocity-field characteristics that a low-field mobility of about  $8000 \text{ cm}^2/\text{V}\cdot\text{s}$  is correctly reproduced for GaAs without the use of any adjustable parameters.

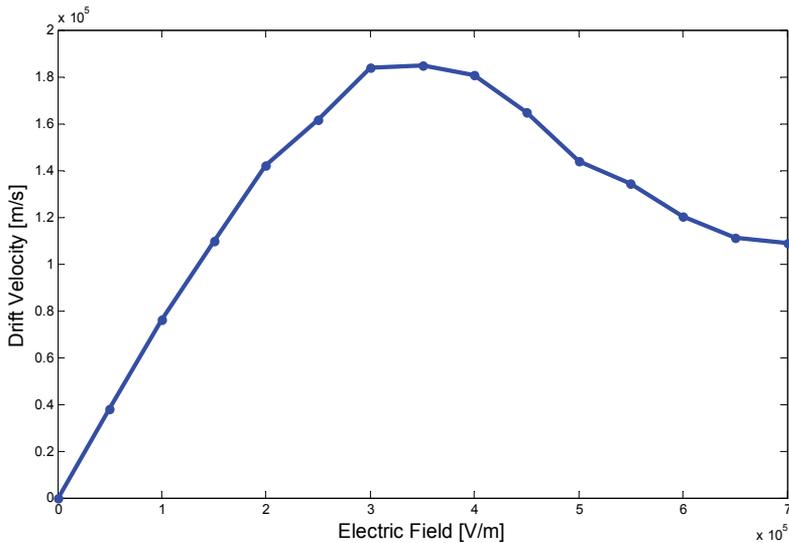


Fig. 17. Steady state drift velocity vs. electric field.

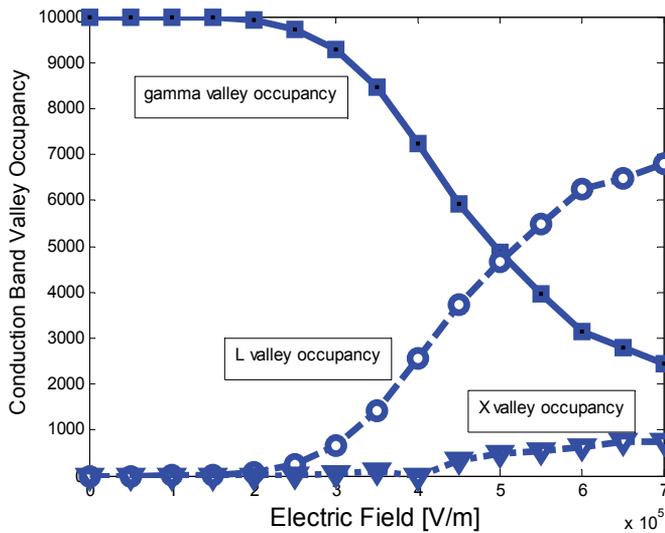


Fig. 18. Different valley occupancy vs. electric field.

At this point, it is advisable to check the energy and wavevector histograms (Figure 19) to ensure that the energy range chosen in the scattering tables is correct or not for the particular maximum electric field strength being considered, which gives the worst case scenario. Since, as already noted, we apply the electric field in the y-direction, for comparative purposes we plot the histograms of the x-component of the wavevector, y-component of the wavevector, and the histogram of the final carrier energy distribution for which a drifted Maxwellian form is evident. Since there is no field applied in the x-direction, we see that the average wavevector in the x-direction is 0. Due to the application of the field in the y-direction, there is a finite positive shift in the y-component of the velocity, which is yet another signature for the displaced Maxwellian form of the energy distribution in the bottom histogram.

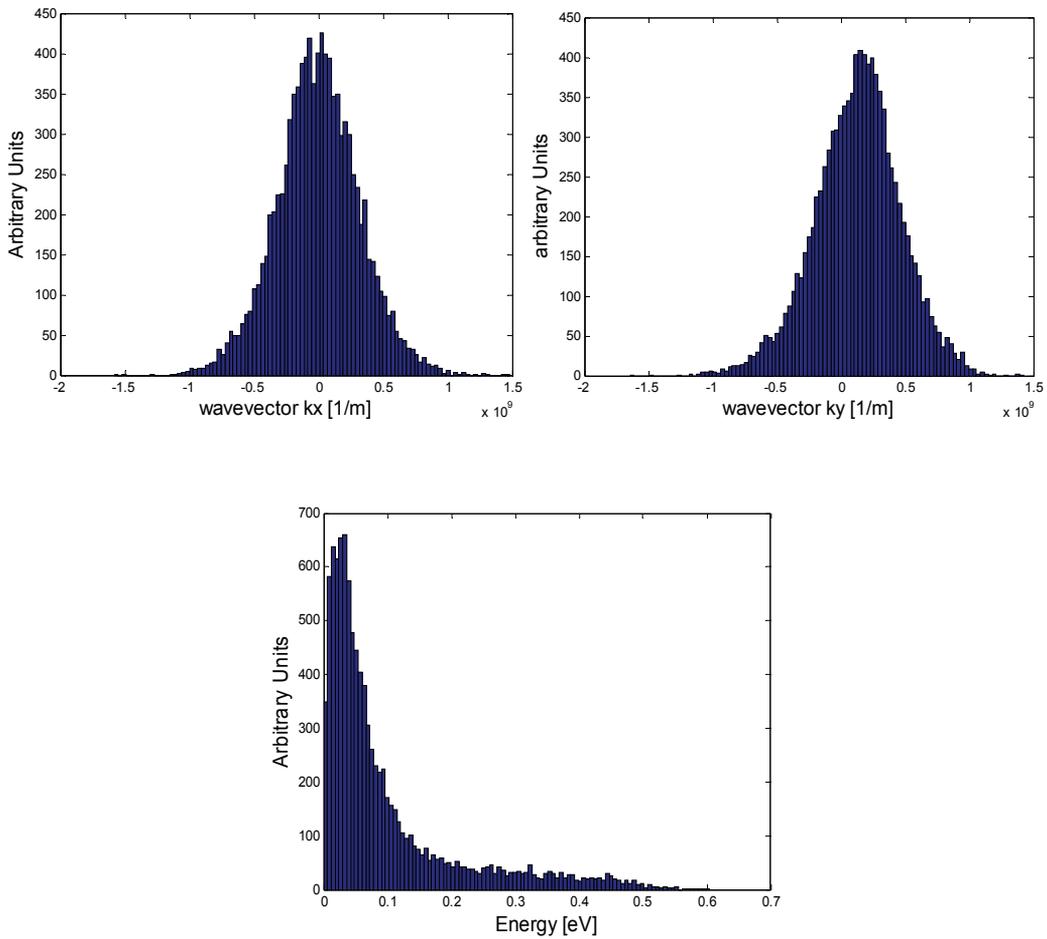


Fig. 19. Top left panel: histogram of the x-component of the wavevector. Top right panel: Histogram of the y-component of the wavevector. Bottom panel: histogram of the carrier energy. Applied electric field is 7kV/cm.

#### 4. Particle-based device simulation

In Section 3.2, we introduced the numerical solution of the BTE using Monte Carlo method. Within a device, both the transport kernel and the field solver are coupled to each other (see Figure 2). The field associated with the potential coming from Poisson's equation is the driving force accelerating particles in the Monte Carlo phase, for example, while the distribution of mobile (both electrons and holes) and fixed charges (e.g. donors and acceptors) provides the source of the electric field in Poisson's equation. Below we give an extensive description of the Monte Carlo particle-based device simulators with emphasis on the particle-mesh coupling.

Within the particle-based EMC method with its time-marching algorithm, Poisson's equation may be decoupled from the BTE over a suitably small time step (typically less than the inverse plasma frequency corresponding to the highest carrier density in the device). Over this time interval, carriers accelerate according to the frozen field profile from the previous time-step solution of Poisson's equation, and then Poisson's equation is solved at the end of the time interval with the frozen configuration of charges arising from the Monte Carlo phase (see discussion in Ref. [45]). Note that Poisson's equation is solved on a mesh, whereas the solution of charge motion using EMC occurs over a continuous range of coordinate space in terms of the particle position. Therefore, a particle-mesh (PM) coupling is needed for both the charge assignment and the force interpolation. The PM coupling is broken into four steps: (1) assign particle charge to the mesh; (2) solve the Poisson equation on the mesh; (3) calculate the mesh-defined forces; and (4) interpolate to find forces on the particle. There are a variety of schemes that can be used for the PM coupling and these are discussed in Section 4.4.

The motion in real space of particles under the influence of electric fields is somewhat more complicated due to the band structure. The velocity of a particle in real space is related to the  $E$ - $\mathbf{k}$  dispersion relation defining the bandstructure as

$$\begin{aligned} \mathbf{v}(t) &= \frac{d\mathbf{r}}{dt} = \frac{1}{\hbar} \nabla_{\mathbf{k}} E(\mathbf{k}(t)) \\ \frac{d\mathbf{k}}{dt} &= \frac{q\mathbf{E}(\mathbf{r})}{\hbar} \end{aligned} \quad (62)$$

where the rate of change of the crystal momentum is related to the local electric field acting on the particle through the acceleration theorem expressed by the second equation. In turn, the change in crystal momentum,  $\mathbf{k}(t)$ , is related to the velocity through the gradient of  $E$  with respect to  $\mathbf{k}$ . If one has to use the full band-structure of the semiconductor, then integration of these equations to find  $\mathbf{r}(t)$  is only possible numerically, using for example a Runge-Kutta algorithm. If a three valley model with parabolic bands is used, then the expression is integrable.

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} = \frac{\hbar\mathbf{k}}{m^*}; \quad \frac{d\mathbf{k}}{dt} = \frac{q\mathbf{E}(\mathbf{r})}{\hbar} \quad (63)$$

Therefore, for a constant electric field in the  $x$  direction, the change in distance along the  $x$  direction is found by integrating twice and is given by equation

$$x(t) = x(0) + v_x(0)t + \frac{qE_x^0 t^2}{2m^*} \quad (64)$$

To simulate the steady-state behavior of a device, the system must be initialized in some initial condition, with the desired potentials applied to the contacts, and then the simulation proceeds in a time stepping manner until steady-state is reached. This process may take several picoseconds of simulation time, and consequently several thousand time-steps based on the usual time increments required for stability. Clearly, the closer the initial state of the system is to the steady state solution, the quicker the convergence. If one is, for example, simulating the first bias point for a transistor simulation, and has no a priori knowledge of the solution, a common starting point for the initial guess is to start out with charge neutrality, i.e. to assign particles randomly according to the doping profile in the device and based on the super-particle charge assignment of the particles, so that initially the system is charge neutral on the average. For two-dimensional device simulation, one should keep in mind that each particle actually represents a rod of charge into the third dimension. Subsequent simulations at the same device at different bias conditions can use the steady state solution at the previous bias point as a good initial guess. After assigning charges randomly in the device structure, charge is then assigned to each mesh point using the NGP or CIC or NEC particle-mesh methods, and Poisson's equation solved. The forces are then interpolated on the grid, and particles are accelerated over the next time step. A flow-chart of a typical Monte Carlo device simulation is shown in Figure 20.

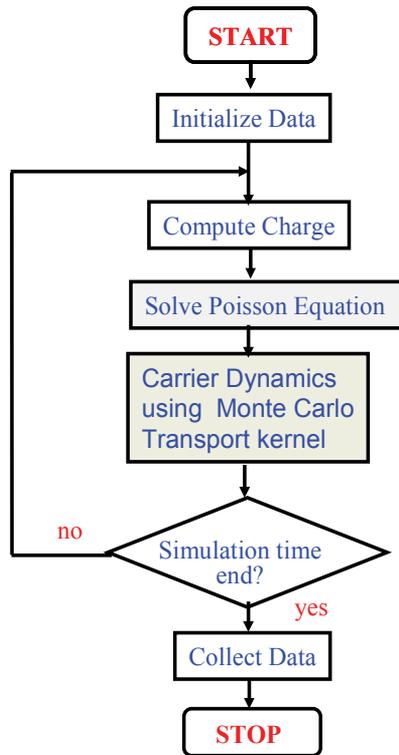


Fig. 20. Flow-chart of a typical particle based device simulation.

As the simulation evolves, charge will flow in and out of the contacts, and depletion regions internal to the device will form until steady state is reached. The charge passing through the contacts at each time step can be tabulated, and a plot of the cumulative charge as a function of time gives the steady-state current. Figure 21 shows the particle distribution in 3D of a MESFET, where the dots indicate the individual simulated particles for two different gate biases. Here, the heavily doped MESFET region (shown by the inner box) is surrounded by semi-insulating GaAs forming the rest of the simulation domain. The upper curve corresponds to no net gate bias (i.e. the gate is positively biased to overcome the built-in potential of the Schottky contact), while the lower curve corresponds to a net negative bias applied to the gate, such that the channel is close to pinch-off. One can see the evident depletion of carriers under the gate under the latter conditions.

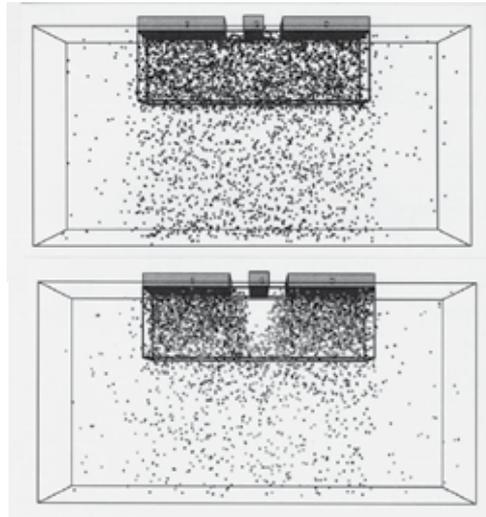


Fig. 21. Example of the particle distribution in a MESFET structure simulated in 3D using an EMC approach. The upper plot is the device with zero gate voltage applied, while the lower is with a negative gate voltage applied, close to pinch-off.

#### 4.1 Calculation of the current

The device output current can be determined using two different yet consistent methods. *First*, by keeping track of the charges entering and exiting each terminal/contact, the net number of charges over a period of the simulation can be used to calculate the terminal current. The net charge crossing a terminal boundary is determined by

$$Q(t) = e(n_{abs}(t) - n_{injec}(t)) + \varepsilon \int E_y(x, t) dy, \quad (65)$$

where  $n_{abs}$  is the number of particles that are absorbed by the contact (exit),  $n_{injec}$  is the number of particles that have been injected at the contact,  $E_y$  is the vertical field at the contact. The second term in Eq. (65) on the right-hand-side is used to account for the displacement current due to the changing field at the contact. Eq. (65) assumes the contact is at the top of the device and that the fields in the  $x$  and  $z$  direction are negligible. The charge  $e$  in Eq. (65) should be multiplied by the particle charge if it is not unity. The slope of

$Q(t)$  versus time gives a measure of the terminal current. In steady state, the current can be found by

$$I = \frac{dQ(t)}{dt} = \frac{e(n_{net})}{\Delta t}, \tag{66}$$

where  $n_{net}$  is the net number of particles exiting the contact over a fixed period of time  $\Delta t$ . The method is quite noisy, due to the discrete nature of the electrons. An example of calculation of the current and keeping the ohmic contacts charge neutral is given Figure 22.

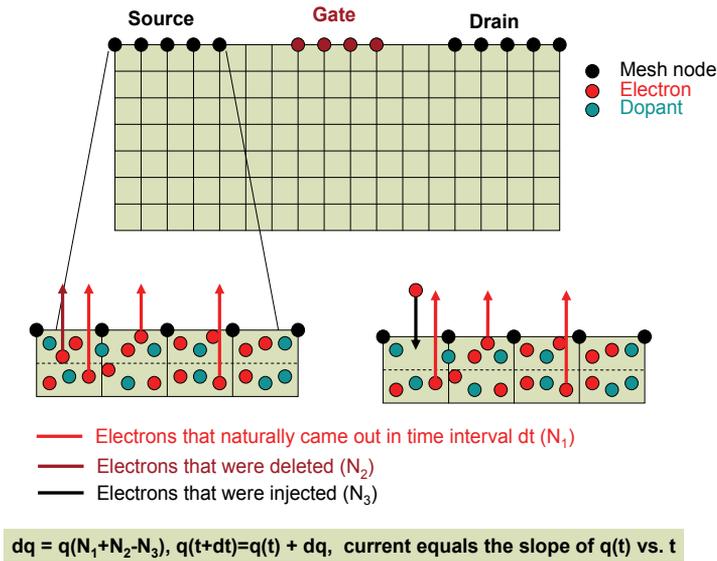


Fig. 22. Keeping charge neutrality at the ohmic contacts and contributions of various terms to the current.

In a *second* method, the sum of the electron velocities in a portion of the channel region of the device is used to calculate the current. The electron current density through a cross-section of the device is given by

$$J = env_d, \tag{67}$$

where  $v_d$  is the average electron drift velocity and  $n$  is the carrier concentration. If there are a total of  $N$  particles in a differential volume,  $dV = dL \cdot dA$ , the current found by integrating Eq. (67) over the cross-sectional area,  $dA$ , is

$$I = \frac{eNv_d}{dL}, \text{ or } I = \frac{e}{dL} \sum_{i=1}^N v_x(i), \tag{68}$$

where  $v_x(i)$  is the velocity along the channel of the  $i^{\text{th}}$  electron. The device is divided into several sections along the  $x$ -axis, and the number of electrons and their corresponding velocity is added for each section after each free-flight. The total  $x$ -velocity in each section is

then averaged over several timesteps to determine the current for that section. Total device current can be determined from the average of several sections, which gives a much smoother result compared to counting terminal charges. By breaking the device into sections, individual currents can be compared to verify that there is conservation of particles (constant current) throughout the device. In addition, sections near the source and drain regions may have a high  $y$ -component in their velocity and should be excluded from the current calculations. Finally, by using several sections in the channel, the average energy and velocity of electrons along the channel can be observed to ensure the proper physical characteristics. The two methods for the calculation of the current are illustrated in Figure 23 on the example of a 50 nm channel length MOSFET device.

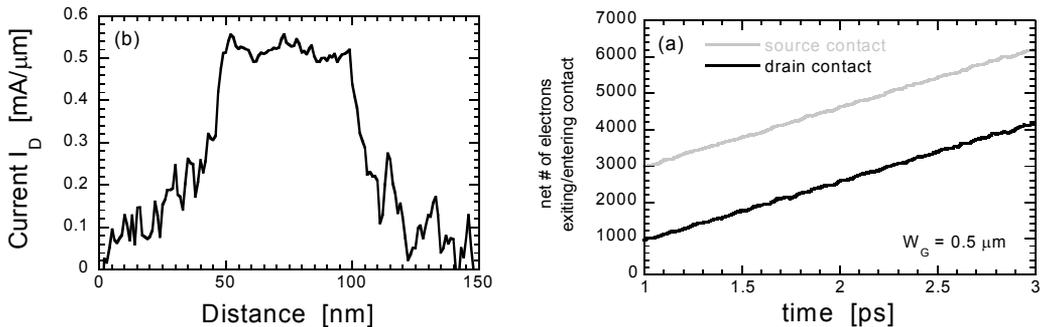


Fig. 23. (Left panel) Net charge entering/exiting the source/drain contact. (Right panel) Average current along the channel. The gate-length of the device being modeled equals 50 nm. We use  $V_G = 1.4$  V and  $V_D = 1$  V in these simulations.

Extrapolating the slope of the curve shown in Figure 23 (left panel), that represents the cumulative electron charge that enters/exits the source/drain contact, leads to source/drain current of 0.5205/0.5193 mA/ $\mu\text{m}$ . When compared with the results shown in Figure 23 (right panel), it is evident that both the current measurement techniques discussed in this section give current values with relative error less than 2 %.

## 4.2 Ohmic contacts

Another issue that has to be addressed in particle-based simulations is the real space boundary conditions for the particle part of the simulation. Reflecting boundary conditions are usually imposed at the artificial boundaries. As far as the ohmic contacts are concerned, they require more careful consideration because electrons crossing the source and drain contact regions contribute to the corresponding terminal current. Commonly employed models for the contacts include [31]:

- Electrons are injected at the opposite contact with the same energy and wavevector  $\mathbf{k}$ . If the source and drain contacts are in the same plane, as in the case of MOSFET simulations, the sign of  $\mathbf{k}$ , normal to the contact will change. This is an unphysical model, however [32].
- Electrons are injected at the opposite contact with a wavevector randomly selected based upon a thermal distribution. This is also an unphysical model.
- Contact regions are considered to be in thermal equilibrium. The total number of electrons in a small region near the contact are kept constant, with the number of

electrons equal to the number of dopant ions in the region. This is a very good model most commonly employed in actual device simulations.

- Another method uses 'reservoirs' of electrons adjacent to the contacts. Electrons naturally diffuse into the contacts from the reservoirs, which are not treated as part of the device during the solution of Poisson's equation. This approach gives results similar to the velocity weighted Maxwellian [31], but at the expense of increased computational time due to the extra electrons simulated. It is an excellent model employed in few most sophisticated particle-based simulators.

There are also several possibilities for the choice of the distribution function — Maxwellian, displaced Maxwellian, and velocity-weighted Maxwellian [33].

### 4.3 Time step

As in the case of solving the Drift-Diffusion, Hydrodynamic or full Maxwell's equations, for a stable Monte Carlo device simulation, one has to choose the appropriate time step,  $\Delta t$ , and the spatial mesh size ( $\Delta x$ ,  $\Delta y$ , and/or  $\Delta z$ ). The time step and the mesh size may correlate to each other in connection with the numerical stability. For example, as discussed in the context of solving Drift-Diffusion simulations, the time step  $\Delta t$  must be related to the plasma frequency

$$\omega_p = \sqrt{\frac{e^2 n}{\epsilon_s m^*}}, \quad (69)$$

where  $n$  is the carrier density. From the viewpoint of the stability criterion,  $\Delta t$  must be much smaller than the inverse plasma frequency. The highest carrier density specified in the device model is used to estimate  $\Delta t$ . If the material is a multi-valley semiconductor, the smallest effective mass to be experienced by the carriers must be used in Eq. (69) as well. In the case of GaAs, with the doping of  $5 \times 10^{17} \text{ cm}^{-3}$ ,  $\omega_p \cong 5 \times 10^{13}$ ; hence,  $\Delta t$  must be smaller than 0.02 ps.

The mesh size for the spatial resolution of the potential is dictated by the charge variations. Hence, one has to choose the mesh size to be smaller than the smallest wavelength of the charge variations. The smallest wavelength is approximately equal to the Debye length, given as

$$\lambda_D = \sqrt{\frac{\epsilon_s k_B T}{e^2 n}}. \quad (70)$$

The highest carrier density specified in the model should be used to estimate  $\lambda_D$  from the stability criterion. The mesh size must be chosen to be smaller than the value given by Eq. (70). In the case of GaAs, with the doping density of  $5 \times 10^{17} \text{ cm}^{-3}$ ,  $\lambda_D \cong 6 \text{ nm}$ .

Based on the discussion above, the time step ( $\Delta t$ ), and the mesh size ( $\Delta x$ ,  $\Delta y$ , and/or  $\Delta z$ ) can be specified separately. However, the  $\Delta t$  chosen must be checked again by calculating the distance  $l_{\max}$ , defined as

$$l_{\max} = \mathbf{v}_{\max} \times \Delta t, \quad (71)$$

where  $v_{max}$  is the maximum carrier velocity that can be approximated by the maximum group velocity of the electrons in the semiconductor (on the order of  $10^8$  cm/s). Therefore, the distance  $l_{max}$  is regarded as the maximum distance the carriers can propagate during  $\Delta t$ . The time step chosen must be small enough so that  $l_{max}$  is smaller than the spatial mesh size chosen using Eq. (71). This is because large  $\Delta t$  chosen may cause substantial change in the charge distribution, while the field distribution in the simulation is only updated every  $\Delta t$ .

**4.4 Particle-mesh (PM) coupling**

As mentioned earlier, the position of charge as described by the EMC algorithm is continuous, whereas Poisson’s equation is solved on a mesh, hence the charge associated with the individual particles must be mapped onto the field mesh in some fashion. The charge assignment and force interpolation schemes usually employed in self-consistent Monte Carlo device simulations are the nearest-grid-point (NGP) and the cloud-in-cell (CIC) schemes [35]. In the NGP scheme, the particle position is mapped into the charge density at the closest grid point to a given particle. This has the advantage of simplicity, but leads to a noisy charge distribution, which may exacerbate numerical instability. Alternately, within the CIC scheme a finite volume is associated with each particle spanning several cells in the mesh, and a fractional portion of the charge per particle is assigned to grid points according to the relative volume of the ‘cloud’ occupying the cell corresponding to the grid point. This method has the advantage of smoothing the charge distribution due to the discrete charges of the particle based method, but may result in an artificial ‘self-force’ acting on the particle, particularly if an inhomogeneous mesh is used. The particle-mesh coupling sequence is presented in Figure 24.

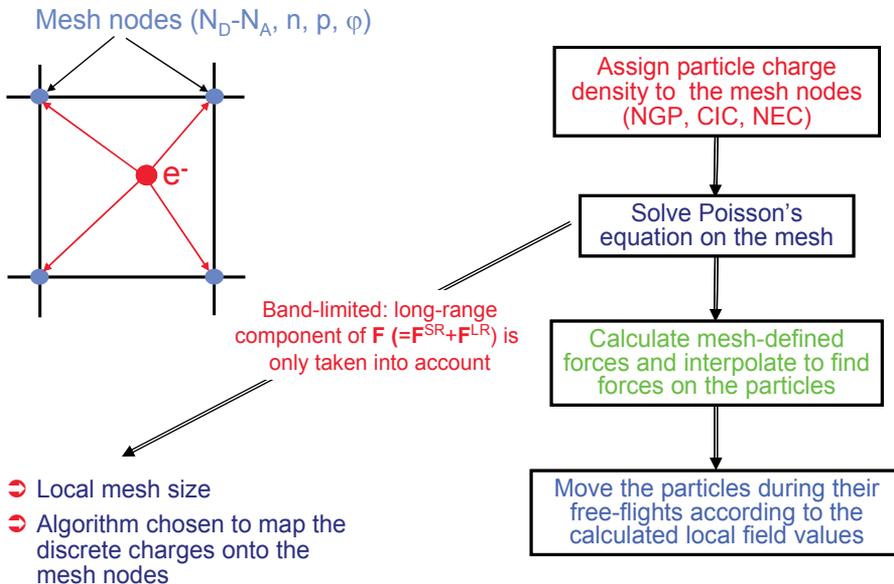


Fig. 24. Particle-mesh coupling sequence.

To better understand the NGP and the CIC scheme, consider a tensor product mesh with mesh lines  $x_i, i = 1, \dots, N_x$  and  $y_j, j = 1, \dots, N_y$ . If the mesh is uniformly spaced in each axis

direction, then  $(x_{l+1} - x_l) = (x_{l+2} - x_{l+1})$ . The permittivities are considered constant within each mesh element and are denoted by  $\varepsilon_{kl}$ ,  $k = 1, \dots, N_x - 1$  and  $l = 1, \dots, N_y - 1$ . Define centered finite-differences of the potential  $\psi$  in the  $x$ - and  $y$ -axis at the midpoints of element edges as follows:

$$\begin{cases} \Delta_{k+\frac{1}{2},l}^x = -\frac{\psi_{k+1,l} - \psi_{k,l}}{x_{k+1} - x_k}, \\ \Delta_{k,l+\frac{1}{2}}^y = -\frac{\psi_{k,l+1} - \psi_{k,l}}{y_{l+1} - y_l}, \end{cases} \quad (72)$$

where the minus sign is included for convenience because the electric field is negative of the gradient of the potential. Consider now a point charge in 2-D located at  $(x, y)$  within an element  $\langle i, j \rangle$ . If the restrictions for the permittivity (P) and the tensor-product meshes with uniform spacing in each direction (M) apply, the standard NGP/CIC schemes in two dimensions can be summarized by the following four steps:

*Charge assignment to the mesh:* The portion of the charge  $\rho_L$  assigned to the element nodes  $(k, l)$  is  $w_{kl}\rho_L$ ,  $k=i, i+1$  and  $l=j, j+1$ , where  $w_{kl}$  are the four charge weights which sum to unity by charge conservation. For the NGP scheme, the node closest to  $(x, y)$  receives a weight  $w_{kl}=1$ , with the remaining three weights set to zero. For the CIC scheme, the weights are  $w_{ij} = w_x w_y$ ,  $w_{i+1,j} = (1 - w_x)w_y$ ,  $w_{i,j+1} = w_x(1 - w_y)$ , and  $w_{i+1,j+1} = (1 - w_x)(1 - w_y)$ ,  $w_x = (x_{i+1} - x)/(x_{i+1} - x_i)$  and  $w_y = (y_{j+1} - y)/(y_{j+1} - y_j)$ .

*Solve the Poisson equation:* The Poisson equation is solved by some of the numerical techniques discussed in Ref. [34].

*Compute forces on the mesh:* The electric field at mesh nodes  $(k, l)$  is computed as:

$$E_{kl}^x = \left( \Delta_{k-\frac{1}{2},l}^x + \Delta_{k+\frac{1}{2},l}^x \right) / 2 \quad \text{and} \quad E_{kl}^y = \left( \Delta_{k,l-\frac{1}{2}}^y + \Delta_{k,l+\frac{1}{2}}^y \right) / 2, \quad \text{for } k = i, i+1 \text{ and } l = j, j+1.$$

*Interpolate to find forces on the charge:* Interpolate the field to position  $(x, y)$  according to  $E^x = \sum_{kl} w_{kl} E_{kl}^x$  and  $E^y = \sum_{kl} w_{kl} E_{kl}^y$ , where  $k = i, i+1$ ,  $l = j, j+1$  and the  $w_{ij}$  are the NGP or CIC weights from step 1.

The requirements (P) and (M) severely limit the scope of devices that may be considered in device simulations using the NGP and the CIC schemes. Laux [35] proposed a new particle-mesh coupling scheme, namely, the nearest-element-center (NEC) scheme, which relaxes the restrictions (P) and (M). The NEC charge assignment/force interpolation scheme attempts to reduce the self-forces and increase the spatial accuracy in the presence of nonuniformly spaced tensor-product meshes and/or spatially-dependent permittivity. In addition, the NEC scheme can be utilized in one axis direction (where local mesh spacing is nonuniform) and the CIC scheme can be utilized in the other (where local mesh spacing is uniform). Such hybrid schemes offer smoother assignment/interpolation on the mesh compared to the pure NEC. The new steps of the pure NEC PM scheme are:

1. *Charge assignment to the mesh:* Divide the line charge  $\rho_L$  equally to the four mesh points of the element  $\langle i, j \rangle$ .
2. *Solve the Poisson equation.*

*Compute forces on the mesh:* Calculate the fields  $\Delta_{i+\frac{1}{2},l}^x, l=j, j+1$ , and  $\Delta_{k,j+\frac{1}{2}}^y, k=i, i+1$ .

*Interpolate to find force on the charge:* Interpolate the field according to the following  $E^x = \left( \Delta_{i+\frac{1}{2},j}^x + \Delta_{i+\frac{1}{2},j+1}^x \right) / 2$  and  $E^y = \left( \Delta_{i,j+\frac{1}{2}}^y + \Delta_{i+1,j+\frac{1}{2}}^y \right) / 2$ .

The NEC designation derives from the appearance, in step (1') of moving the charge to the center of its element and applying a CIC-like assignment scheme. The NEC scheme involves only one mesh element and its four nodal values of potential. This locality makes the method well-suited to non-uniform mesh spacing and spatially-varying permittivity. The interpolation and error properties of the NEC scheme are similar to the NGP scheme.

#### 4.5 Higher order effects

Multi-particle effects relate to the interaction between particles in the system, which is a nonlinear effect when viewed in the context of the BTE, due to the dependence of such effects on the single particle distribution function itself. Most algorithms developed to deal with such effects essentially linearize the BTE by using the previous value of the distribution function to determine the time evolution of a particle over the successive time-step. Multi-carrier effects may range from simple consideration of the Pauli exclusion principle (which depends on the exact occupancy of states in the system), to single particle and collective excitations in the system. Inclusion of carrier-carrier interactions in Monte Carlo simulation has been an active area of research for quite some time and is briefly discussed below. Another carrier-carrier effect, that is of considerable importance when estimating leakage currents in MOSFET devices, is impact ionization, which is a pure generation process involving three particles (two electrons and a hole or two holes and an electron). The latter is also discussed below.

##### 4.5.1 Pauli exclusion principle

The Pauli exclusion principle requires that the bare scattering rate be modified by a factor  $1 - f_m(\mathbf{k}')$  in the collision integral of the BTE, where  $f_m(\mathbf{k}')$  is the one-particle distribution function for the state  $\mathbf{k}'$  in band (subband)  $m$  after scattering. Since the net scattering rate including the Pauli exclusion principle is always less than the bare scattering rate, a self-scattering rejection technique may be used in the Monte Carlo simulation as proposed by Bosi and Jacoboni [36] for one particle simulation and extended by Lugli and Ferry [37] for EMC. In the self-scattering rejection algorithm, an additional random number  $r$  is generated (between 0 and 1), and this number is compared to  $f_m(\mathbf{k}')$ , the occupancy of the final state (which is also between 0 and 1 when properly normalized for the numerical  $\mathbf{k}$ -space discretization). If  $r$  is greater than  $f_m(\mathbf{k}')$ , the scattering is accepted and the particle's momentum and energy are changed. If this condition is not satisfied, the scattering is rejected, and the process is treated as a self-scattering event with no change of energy or momentum after scattering. Through this algorithm, it is clear that no scattering occurs if the final state is completely full.

##### 4.5.2 Carrier-carrier interactions

Carrier-carrier interactions, apart from degeneracy effects, may be treated as a scattering process within the Monte Carlo algorithm on the same footing as other mechanisms. In the simplest case of bulk electrons in a single parabolic conduction band, the process may be

treated as a binary collision where the scattering rate for a particle of wavevector  $\mathbf{k}_0$  due to all the other particles in the ensemble is given by [38]

$$\Gamma_{ee}(\mathbf{k}_0) = \frac{nm_n e^4}{4\pi\hbar^3 \varepsilon^2 \beta^2} \int d\mathbf{k} f(\mathbf{k}) \frac{|\mathbf{k} - \mathbf{k}_0|}{(|\mathbf{k} - \mathbf{k}_0|^2 + \beta^2)}, \quad (73)$$

where  $f(\mathbf{k})$  is the one-particle distribution function (normalized to unity),  $\varepsilon$  is the permittivity,  $n$  is the electron density, and  $\beta$  is the screening constant. In deriving Eq. (73), one assumes that the two particles interact through a statically screened Coulomb interaction, which ignores the energy exchange between particles in the screening which in itself is a dynamic, frequency-dependent effect. Similar forms have been derived for electrons in 2D [39,40] and 1D [41], where carrier-carrier scattering leads to inter-subband as well as intra-subband transitions. Since the scattering rate in Eq. (73) depends on the distribution function of all the other particles in the system, this process represents a nonlinear term as discussed earlier. One method is to tabulate  $f(\mathbf{k})$  on a discrete grid, as is done for the Pauli principle, and then numerically integrate Eq. (73) at each time step. An alternate method is to use a self-scattering rejection technique [42], where the integrand excluding  $f(\mathbf{k})$  is replaced by its maximum value and taken outside the integral over  $\mathbf{k}$ . The integral over  $f(\mathbf{k})$  is just unity, giving an analytic form used to generate the free flight. Then, the self-scattering rejection technique is used when the final state is chosen to correct for the exact scattering rate compared to this artificial maximum rate, similar to the algorithm used for the Pauli principle.

The treatment of intercarrier interactions as binary collisions above neglects scattering by collective excitations such as plasmons or coupled plasmon-phonon modes. These effects may have a strong influence on carrier relaxation, particularly at high carrier density. One approach is to make a separation of the collective and single particle spectrum of the interacting many-body Hamiltonian, and treat them separately, i.e. as binary collisions for the single particle excitations, and as electron-plasmon scattering for the collective modes [43]. Another approach is to calculate the dielectric response within the random phase approximation, and associate the damping given by the imaginary part of the inverse dielectric function with the electron lifetime [44].

A semiclassical approach to carrier-carrier interaction, which is fully compatible with the Monte Carlo algorithm, is the use of Molecular Dynamics [45], in which carrier-carrier interaction is treated continuously in real space during the free-flight phase through the Coulomb force of all the particles. A very small time step is required when using Molecular Dynamics to account for the dynamic distribution of the system. A time step on the order of 0.5 fs is often sufficiently small for this purpose. The small time step assures that the forces acting on the particles during the time of flight are essentially constant, that is  $f(t) \cong f(t + \Delta t)$ , where  $f(t)$  is the single particle distribution function.

Using Newtonian kinematics, we can write the real space trajectories of each particle as

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}\Delta t + \frac{1}{2} \frac{\mathbf{F}(t)}{m} \Delta t^2, \quad (74)$$

and

$$\mathbf{v}(t + \Delta t) = \mathbf{v}(t) + \frac{\mathbf{F}(t)}{m} \Delta t. \quad (75)$$

Here,  $\mathbf{F}(t)$  is the force arising from the applied field as well as that of the Coulomb interactions. We can write  $\mathbf{F}(t)$  as

$$\mathbf{F}(t) = q \left[ \mathbf{E} - \sum_i \nabla \varphi(\mathbf{r}_i(t)) \right], \quad (76)$$

where  $q\mathbf{E}$  is the force due to the applied field and the summation is the interactive force due to all particles separated by distance  $\mathbf{r}_i$ , with  $\varphi(\mathbf{r}_i)$  the electrostatic potential. As in Monte Carlo simulation, one has to simulate a finite number of particles due to practical computational limitations on execution time. In real space, this finite number of particles corresponds to a particular simulation volume given a certain density of carriers,  $V = N/n$ , where  $n$  is the density. Since the carriers can move in and out of this volume, and since the Coulomb interaction is a long-range force, one must account for the region outside  $V$  by periodically replicating the simulated system. The contributions due to the periodic replication of the particles inside  $V$  in cells outside has a closed form solution in the form of an Ewald sum [46], which gives a linear as well as  $1/r^2$  contribution to the force. The equation for the total force in the Molecular Dynamics technique then becomes

$$\mathbf{F} = \frac{-e^2}{4\pi\epsilon} \sum_i^N \left( \frac{1}{\mathbf{r}_i^2} \mathbf{a}_i + \frac{2\pi}{3V} \mathbf{r}_i \right). \quad (77)$$

The above equation is easily incorporated in the standard Monte Carlo simulation discussed up to this point. At every time step the forces on each particle due to all the other particles in the system are calculated from Eq. (77). From the forces, an interactive electric field is obtained which is added to the external electric field of the system to couple the Molecular Dynamics to the Monte Carlo.

The inclusion of the carrier-carrier interactions in the context of particle-based device simulations is discussed in Ref. [47]. The main difficulty in treating this interaction term in device simulations arises from the fact that the long-range portion of the carrier-carrier interaction is included via the numerical solution of the quasi-static Poisson equation. Under these circumstances, special care has to be taken when incorporating the short-range portion of this interaction term to prevent double counting of the force.

#### 4.5.3 Band to Band impact ionization

Another carrier-carrier scattering process is that of impact ionization, in which an energetic electron (or hole) has sufficient kinetic energy to create an electron-hole pair. Impact ionization therefore leads to the process of carrier multiplication. This process is critical for example in the avalanche breakdown of semiconductor junctions, and is a detrimental effect in short channel MOS devices in terms of excess substrate current and decreased reliability.

The ionization rate of valence electrons by energetic conduction band electrons is usually described by Fermi's rule in which a screened Coulomb interaction is assumed between the two particles, where screening is described by an appropriate dielectric function such as that proposed by Levine and Louie [48]. In general, the impact ionization rate should be a function of the wavevector of the incident electron, hence of the direction of an electric field in the crystal, although there is still some debate as to the experimental and theoretical evidence. More simply, the energy dependent rate (averaged over all wavevectors on a constant energy shell) may be expressed analytically in the power law form

$$\Gamma_{ii}(E) = P[E - E_{th}]^a, \quad (78)$$

where  $E_{th}$  is the threshold energy for the process to occur, which is determined by momentum and energy conservation considerations, but minimally is the bandgap of the material itself.  $P$  and  $a$  are parameters which may be fit to more sophisticated models. The Keldysh formula [49] is derived by expanding the matrix element for scattering close to threshold, which gives  $a=2$ , and the constant  $P=C/E_{th}^2$ , with  $C=1.19 \times 10^{14}/s$  and assuming a parabolic band approximation,

$$E_{th} = \frac{3 - 2m_v/m_c}{1 - m_v/m_c} E_g, \quad (79)$$

where  $m_v$  and  $m_c$  are the effective masses of the valence and conduction band respectively, and  $E_g$  is the bandgap. More complete full-bandstructure calculations of the impact ionization rate have been reported for Si [50,51], GaAs [51,52] and wide bandgap materials [53], which are fairly well fit using using power law model.

Within the ensemble Monte Carlo method, the scattering rate given by Eq. (78) is used to generate the free flight time. The state after scattering of the initial electron plus the additional electron and hole must satisfy both energy and momentum conservation within the Fermi rule model, which is somewhat complicated unless simple parabolic band approximations are made.

## 5. What is the future for Monte Carlo device simulations?

The drift-diffusion solvers are applicable in situations in which the bias conditions and the device geometry are such that electric fields are relatively low and velocity saturation model is applicable. Situations in which drift-diffusion models are applicable are the silicon based power MOSFET devices, bipolar junction transistors, light emitting diodes (that are used more and more in solid state lightning) and crystalline solar cells, just to name a few. In some of these devices, such as power transistors and LEDs used for solid state lighting, it is of paramount importance to incorporate self-heating models within the drift-diffusion framework. The accuracy of simple heating models in conjunction with drift-diffusion models is to some degree questionable so that it is in many circumstances justifiable to use particle-based device simulators with more exact self-heating models.

On the other hand, the hydrodynamic models do not suffer from the limitations of the drift-diffusion approaches and the incorporation of the additional energy balance equation allows one to include velocity overshoot in the model. Velocity overshoot and non-stationary transport are key features of conventional MOSFET devices with gate lengths of 200 nm and below. However, as it was explained in Section 1.3 of this book chapter, the magnitude of the velocity overshoot observed via simulations depends strongly upon the choice of the energy relaxation time, mostly in sub-100 nm channel length devices. This, in turn, affects the magnitude of the drain current. The reason for such drastic differences in the results when different energy relaxation times are used is the fact that the energy relaxation time is material as well as device geometry dependent parameter. So, to calculate better estimates for the velocity overshoot, higher moments of the Boltzmann transport equation are needed. These, in turn involve parameters that are more and more ambiguous on the expense of increased computational cost and when the computational cost of hydrodynamic models exceeds the one of particle-based device simulators, there is no point in using moment

methods. In these circumstances the direct solution of the Boltzmann Transport Equation via the Monte Carlo method becomes a method of choice. Thus, we might conclude that it is advisable to use particle-based device simulators when nano-scale devices are concerned.

But how far down in the scaling can we go? Particle-based device simulators capture on an equal footing ballistic and diffusive transport, so if the ballisticity factor in the device increases, there is no problem that ballistic transport is effectively captured with particle-based device simulators. Quantum mechanical size quantization effects can also be captured by solving in slices the corresponding 1D or 2D Schrödinger equation if one is concerned with conventional or fully-depleted SOI MOSFETs, or nanowire transistors, respectively. What can not be captured with particle-based device simulators is if there are local strains and stresses in the ultra-nano-scale devices, but that can also be cured via coupling of Monte-Carlo device simulators with atomistic models for band-structure calculation.

In summary, Monte Carlo device simulators are a powerful tool for modeling devices ranging from the nano-scale regime to the microscale regime. What can not be modeled with particle-based device simulators are resonant tunneling diodes in which quantum interference effects dominate the device behavior. Efforts have been made along this direction as well, but the inclusion of the quantum-mechanical phase alongside with well defined particle trajectory still remains open field of research.

## 6. References

- [1] David K. Ferry and Stephen M. Goodnick, *Transport in Nanostructures* (Cambridge Studies in Semiconductor Physics and Microelectronic Engineering, 1997).
- [2] Vasileska and S. M. Goodnick, *Materials Science and Engineering, Reports: A Review: Journal*, R38, no. 5, 181 (2002).
- [3] S. M. Goodnick and D. Vasileska, *Encyclopedia of Materials: Science and Technology*, Vol. 2, Ed. By K. H. J. Buschow, R. W. Cahn, M. C. Flemings, E. J. Kramer and S. Mahajan, Elsevier, New York, 1456, (2001).
- [4] D. Vasileska and S. M. Goodnick, *Computational Electronics* (Morgan and Claypool, 2006).
- [5] A. Schütz, S. Selberherr, H. Pötzl, *Solid-State Electronics*, Vol. 25, 177 (1982).
- [6] P. Antognetti and G. Massobrio, *Semiconductor Device Modeling with SPICE* (McGraw-Hill, New York, 1988).
- [7] M. Shur, *Physics of Semiconductor Devices* (Prentice Hall Series in Solid State Physical Electronics).
- [8] D. L. Scharfetter and D. L. Gummel, *IEEE Transaction on Electron Devices*, Vol. ED-16, 64 (1969).
- [9] K. Bløtekjær, *IEEE Trans. Electron Dev.*, Vol. 17, 38 (1970).c
- [10] M. V. Fischetti and S. E. Laux, "Monte Carlo Simulation of Submicron Si MOSFETs", *Simulation of Semiconductor Devices and Processes*, vol. 3, G. Baccarani and M. Rudan Eds. (Technoprint, Bologna, 1988), 349.
- [11] L. V. Keldysh, *Sov. Phys. – JETP*, Vol. 20, 1018 (1965).
- [12] A. L. Fetter, J. D. Walecka, *Quantum Theory of Many-Particle Systems* (McGraw-Hill 1971).
- [13] G. D. Mahan, *Many-Particle Physics* (Kluwer Academic/Plenum Publishers, New York, 2000).
- [14] R. Lake, G. Klimeck, R.C. Bowen, and D. Jovanovic, *J. Appl. Phys.*, Vol. 81, 7845 (1997)
- [15] G. Baccarani, M. Wordeman, *Solid State Electron.*, Vol. 28 , 407 (1985).
- [16] S. Cordier, *Math. Mod. Meth. Appl. Sci.*, Vol. 4, 625 (1994).
- [17] K. Tomizawa, *Numerical Simulation of Submicron Semiconductor Devices* (The Artech House Materials Science Library).

- [18] H. K. Gummel, *IEEE Transactions on Electron Devices*, Vol. 11, 455 (1964).
- [19] T. M. Apostol, *Calculus, Vol. II, Multi-Variable Calculus and Linear Algebra* (Blaisdell, Waltham, MA, 1969) ch. 1.
- [20] R. Straton, *Phys. Rev.*, Vol. 126, 2002 (1962).
- [21] T. Grasser, T.-W. Tang, H. Kosina, and S. Selberherr, *Proceedings of the IEEE*, Vol. 91, 251 (2003).
- [22] M.A. Stettler, M.A. Alam, and M.S. Lundstrom, *Proceedings of the NUPAD Conference*, 97 (1992).
- [23] [www.silvaco.com](http://www.silvaco.com)
- [24] C. Jacoboni and L. Reggiani, *Rev. Mod. Phys.*, Vol. 55, 645 (1983).
- [25] C. Jacoboni and P. Lugli, *The Monte Carlo Method for Semiconductor Device Simulation*, Springer-Verlag, Vienna (1989).
- [26] K. Hess, *Monte Carlo Device Simulation: Full Band and Beyond*, (Kluwer Academic Publishing, Boston, 1991).
- [27] M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods*, (Wiley, New York, 1986).
- [28] D. K. Ferry, *Semiconductors*, (Macmillan, New York, 1991).
- [29] H. D. Rees, *J. Phys. Chem. Solids*, Vol. 30, 643 (1969).
- [30] R. M. Yorston, *J. Comp. Phys.*, Vol. 64, 177 (1986).
- [31] T. Gonzalez and D. Pardo, *Solid State Electron.*, 39 (1996) 555.
- [32] P. A. Blakey, S. S. Cherensky and P. Sumer, *Physics of Submicron Structures*, Plenum Press, New York, (1984).
- [33] T. Gonzalez and D. Pardo, *Solid-State Electron.*, 39, 555 (1996).
- [34] D. Vasileska and S.M. Goodnick, "Computational Electronics", *Morgan & Claypool*, 2006
- [35] S. E. Laux, *IEEE Trans. Comp.-Aided Des. Int. Circ. Sys.*, 15, 1266 (1996).
- [36] S. Bosi S and C. Jacoboni, *J. Phys. C*, 9, 315 (1976).
- [37] P. Lugli and D. K. Ferry, *IEEE Trans. Elec. Dev.*, 32, 2431 (1985).
- [38] N. Takenaka, M. Inoue and Y. Inuishi, *J. Phys. Soc. Jap.*, 47, 861 (1979).
- [39] S. M. Goodnick and P. Lugli, *Phys. Rev. B*, 37 (1988) 2578.
- [40] M. Moško, A. Mošková and V. Cambel, *Phys. Rev. B*, 51, 16860 (1995).
- [41] L. Rota, F. Rossi, S. M. Goodnick, P. Lugli, E. Molinari and W. Porod, *Phys. Rev. B*, 47,1632 (1993).
- [42] R. Brunetti, C. Jacoboni, A. Matulionis and V. Dienys, *Physica B&C*, 134, 369 (1985).
- [43] P. Lugli and D. K. Ferry, *Phys. Rev. Lett.*, 56, 1295 (1986).
- [44] J. F. Young and P. J. Kelly, *Phys. Rev. B*, 47, 6316 (1993).
- [45] R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles*, Institute of Physics Publishing, Bristol, (1988).
- [46] D. J. Adams and G. S. Dubey, *J. Comp. Phys.*, 72, 156 (1987).
- [47] D. Vasileska, H.R. Khan, S.S. Ahmed, "Modeling Coulomb effects in nanoscale devices", *Journal of Computational and Theoretical Nanoscience*, Volume 5, Number 9, September 2008, pp. 1793-1827(35).
- [48] Z. H. Levine, and S. G. Louie, *Phys. Rev. B*, 25, 6310 (1982).
- [49] L. V. Keldysh, *Zh. Eksp. Teor. Fiz.*, 37, 713 (1959).
- [50] N. Sano and A. Yoshii, *Phys. Rev. B*, 45, 4171 (1992).
- [51] M. Stobbe, R. Redmer and W. Schattke, *Phys. Rev. B*, 47, 4494 (1994).
- [52] Y. Wang and K. Brennan, *J. Appl. Phys.*, 71, 2736 (1992).
- [53] M. Reigrotzki, R. Redmer, N. Fitzer, S. M. Goodnick, M. Dür, and W. Schattke, *J. Appl. Phys.*, 86, 4458, (1999).

# Wang-Landau Algorithm and its Implementation for the Determination of Joint Density of States in Continuous Spin Models

Soumen Kumar Roy<sup>1</sup>, Kisor Mukhopadhyay<sup>2</sup>, Nababrata Ghoshal<sup>3</sup> and Shyamal Bhar<sup>4</sup>

<sup>1</sup>*Department of Physics, Jadavpur University, Kolkata - 700032*

<sup>2</sup>*Department of Physics, Sundarban Mahavidyalaya, Kakdwip, South 24 Parganas, West Bengal*

<sup>3</sup>*Department of Physics, Mahishadal Raj College, Mahishadal, Purba Medinipur, West Bengal*

<sup>4</sup>*Department of Physics, Vidyasagar College for Women, Kolkata-700006 INDIA*

## 1. Introduction

The Metropolis algorithm (Metropolis et al., 1953) developed more than half a century ago has been extensively used to simulate a wide variety of problems of interest in statistical physics. The method is based on importance sampling and directly generates the equilibrium configurations in the canonical distribution of a system. Thus Monte Carlo (MC) simulations using the Metropolis algorithm are carried out at a number of closely spaced temperatures and the observables are calculated by taking the configurational averages over the states generated by the algorithm. The method does not directly lead to an evaluation of the entropy or density of states of a system.

Although the Metropolis algorithm, which is rather simple to implement in a computer, is still being widely used it suffers from a number of notable drawbacks. In case of a system having a first order phase transition, the random walker in the Metropolis algorithm gets trapped in metastable states, thus leading to long runs which mostly result in inaccurate results. Also in the case of systems exhibiting a second order phase transition, critical slowing down reduces the efficiency of the algorithm to a large extent. To overcome such problems, for a spin system other algorithms have been developed and the Swendsen-Wang algorithm (Swendsen & Wang, 1987) was an important step in this direction. Wolff cluster algorithm (Wolff, 1989) which was developed later has proved to be very important in reducing critical slowing down in continuous spin systems. Using these algorithms, accurate estimation of thermodynamic quantities has become possible using the histogram reweighting technique which was developed by Ferrenberg and Swendsen in 1988 (Ferrenberg & Swendsen, 1988). This works equally well for systems exhibiting both first order and second order phase transitions.

The above mentioned algorithms and the data analysis techniques depend on the generation of configurations of a system at different temperatures. For a few decades after standard

Monte Carlo simulation began following the publication of the Metropolis approach, there existed no method which directly or indirectly evaluates the partition function of a system. A knowledge of the partition function (as a function of temperature) would lead to a straight forward evaluation of most thermodynamic quantities of interest and would prove to be very attractive.

The partition function of a system can easily be evaluated if the density of states  $\Omega(E)$  of a system for a given energy  $E$  is known. The density of states is a quantity independent of temperature and depends on the number of configurations which a system possesses for a given energy. The partition function can be written as

$$Z(\beta) = \sum_E \Omega(E) e^{-\beta E} \quad (1)$$

where  $\beta = 1/K_B T$ ,  $T$  being the temperature and  $K_B$  the Boltzmann factor. It may be noted that  $\Omega(E)$ , although called the density of states (DOS) is actually the number of states or simply the degeneracy factor of a state of the system with energy  $E$ . The above relationship therefore shows that once we have a knowledge of  $\Omega(E)$ , the partition function at any temperature can easily be evaluated simply by Boltzmann reweighting. Without going into the history of development of Monte Carlo techniques which leads to the determination of the density of states (since such material may be found in books like Newman and Barkema (Newman & Barkema, 1999) or Binder and Landau (Landau & Binder, 2000)) we name two techniques which have been proposed in the past one decades or so. These are the Transition Matrix Monte Carlo method (TMCMC) and the Wang-Landau algorithm (WL). The TMCMC method, first proposed by de Oliveria et al. in 1996, makes use of a book keeping of all transitions between the microstates of a system as a random walk is performed in the energy space and the transition probabilities are used to give an accurate estimation of the density of states.

The subject of this chapter is to discuss the other method, the Wang-Landau algorithm, first proposed in 2001 (Wang & Landau, 2001). This has since attracted a wide attention of researchers and have been applied to a large variety of systems. In the original work Wang and Landau applied this method to simulate discrete spin systems showing first order (Pott's model (Wu, 1982)) as well as second order (two-dimensional Ising model (Onsager, 1944)) phase transitions. Also a system having rough energy landscape, like the Edwards-Anderson model (Edwards & Anderson, 1975) of three-dimensional spin glass, which is rather difficult to be simulated using conventional Monte Carlo algorithms, was simulated using the WL algorithm.

The WL algorithm, to be discussed in detail in the following section, is an iterative scheme where during each iteration an histogram of the energy distribution is generated during a random walk in the energy space. Starting from no knowledge of the density of states of a system, one gradually builds up the profile of the density of states by forcing the random walker to make more visits to those energy regions where the density of states is smaller. The method is based on the generation of a flat histogram for each iteration. A small number, called the modification factor is used to build up the DOS profile during each iteration and as the process goes on this modification factor is made smaller and smaller thus making finer adjustments to the value of the DOS.

In the original work of Wang and Landau application of the algorithm only to discrete spin systems were discussed. Also the random walk performed was one-dimensional – only in the energy space. The statistical average of an observable which is directly related to energy is thus given by

$$\langle f(E, \beta) \rangle = \frac{\sum_E f(E) \Omega(E) e^{-\beta E}}{\sum_E \Omega(E) e^{-\beta E}} \quad (2)$$

However if one intends to determine quantities ( $\phi$ ) like the order parameter, correlation function etc., which are not directly related to energy, a two dimensional random walk in the  $E - \phi$  space needs to be performed. Consequently the DOS will also be a function of two variables – energy  $E$  and the other observable  $\phi$ . This is generally called the joint density of states (JDOS) and is denoted by  $\Omega(E, \phi)$ . The ensemble average of any function of  $\phi$  is obtained from the relation:

$$\langle f(\phi, \beta) \rangle = \frac{\sum_E \sum_\phi f(\phi) \Omega(E, \phi) e^{-\beta E}}{\sum_E \sum_\phi \Omega(E, \phi) e^{-\beta E}} \quad (3)$$

Apart from the necessity of generating JDOS, another modification of the original WL algorithm needs to be done even when simulating spin systems. Systems like the XY-, Heisenberg- and the Lebwohl-Lasher (LL) (Lebwohl & Lasher, 1972) models are examples of lattice spin models which have a continuous energy spectrum. It is possible to generate both DOS and JDOS in such system using WL algorithm. The basic algorithm remains the same but one discretizes the energy values by dividing the energy range of interest into narrow bins. In the case of a two-dimensional random walk both the  $E$ -space and  $\phi$ - space are discretized. Thus in the most general case we need to handle the generation of JDOS in a continuous lattice-spin model. This is the subject matter of this chapter (Mukhopadhyay et al., 2008). One can of course apply the WL algorithm to systems other than lattice-spin models but we shall not go into this. With a knowledge of simulating JDOS in continuous spin systems it is straight forward to apply the algorithm to simulate any other system of interest.

In the following section we discuss the WL algorithm and how one can implement this for a discrete spin system. Next we discuss the one-dimensional LL model and this is followed by a section where the technique for handling a continuous model is discussed. This model is chosen because it is the only continuous spin model (although not exhibiting a phase transition) which has an exact solution (Vuillermot & Romerio, 1973; 1975) and thus will allow us to check the accuracy of the results of our simulation. We have also applied the model to a two-dimensional XY model – known to exhibit the Kosterlitz-Thouless (Kosterlitz & Thouless, 1973; Kosterlitz, 1974) phase transition mediated by topological defects. We have compared our results obtained from one-dimensional random walk for the XY model with those obtained from the standard Metropolis algorithm to check the accuracy of our simulation. Results obtained from the simulated JDOS are also presented for this model.

We conclude this section with an important comment, nowhere in their original paper Wang and Landau discussed why the DOS obtained using their algorithm should be the true value of the DOS of a system. The question of convergence of the DOS obtained by the iterative process is also skipped. However, some authors have dealt with this question in more recent papers and we, in a concluding section, will discuss briefly this aspect of the WL algorithm.

## 2. The Wang Landau algorithm

Using an iterative scheme the WL method determines the density of states  $\Omega(E)$  of a system as a function of energy  $E$ . In this method random walk in discrete energy space of a spin system is performed by flipping spins in a random manner. A random walker without any bias tends to visit regions of energy where  $\Omega(E)$  is greater. In order to produce a "flat

histogram", the WL algorithm visits states with a probability which is inversely proportional to  $\Omega(E)$  (instead of sampling with the Boltzmann weights  $e^{-\beta E}$  used in the conventional Metropolis like algorithms). However, since the density of states  $\Omega(E)$  is not known *a priori* all  $\Omega(E)$ s are set equal to a common constant value, say 1 at the beginning of a simulation. At every step of the random walk  $\Omega(E)$  is modified by a multiplicative factor  $f > 1$  and the updated  $\Omega(E)$  is used for the next step of random walk. The modification factor  $f$  is controlled carefully in the following iterations and finally when  $f \approx 1$  the density of states  $\Omega(E)$  converges very close to its true value. The accuracy of the estimated density of states depends on many factors involved in the implementation part of the algorithm such as the final value of the modification factor, flatness criterion, system size etc. Since density of states  $\Omega(E)$  is a large quantity, it is convenient to use  $\ln[\Omega(E)]$  in simulation and we denote it by  $g(E)$ .

The steps of the Wang Landau algorithm in a lattice-spin system having discrete energy values goes like this:

1. Set  $g(E) = 0$  and  $h(E) = 0$  for all  $E$ , here  $h(E)$  is the histogram count. Also choose  $\ln f = 1$  (the initial value of the modification factor may be  $f = e^1$  which is suitable for a faster estimation of  $\Omega(E)$ ).
2. Pick a single spin at random and flip it.
3. Whether the move is accepted or not is decided by the transition probability

$$p_{i \rightarrow j} = \min \left[ \frac{\Omega(E_i)}{\Omega(E_j)}, 1 \right]$$

or,

$$p_{i \rightarrow j} = \min \left[ \exp[g(E_i) - g(E_j)], 1 \right],$$

where  $E_i$  and  $E_j$  are respectively the system energies before and after a spin is flipped. Generate a random number  $r$  such that its value lies between 0 and 1. If  $r$  is less than  $p_{i \rightarrow j}$  flip the spin. Otherwise restore the previous configuration.

4. If the proposed state  $j$  is accepted, make

$$g(E_j) = g(E_j) + \ln f, \quad h(E_j) = h(E_j) + 1$$

otherwise make

$$g(E_i) = g(E_i) + \ln f, \quad h(E_i) = h(E_i) + 1$$

5. Repeat steps 2, 3, and 4 for about  $10^4$  MC sweeps and check the flatness of the histogram. One Monte Carlo sweep consists of a sequence of moves whose number is equal to the number of spins in the system. Since it is not possible to obtain a perfectly flat (100%) histogram, the flatness check is performed by calculating the average histogram  $\langle h(E) \rangle$  averaged over  $E$  and verifying the condition  $\frac{h(E)}{\langle h(E) \rangle} \geq x$  for all  $E$  where  $x$  can be chosen according to the size and complexity of the system (for example  $x$  may be chosen to be 0.9). When the flatness criterion is satisfied for all  $E$  we say that one iteration is completed.
6. In the next iteration reset the histogram counts to zero, reduce the modification factor as  $f \rightarrow \sqrt{f}$  i.e.  $\ln f \rightarrow \ln f / 2$ . The steps (from 2 to 5) are repeated until  $\ln f$  becomes smaller than a very small predefined value say  $\sim 10^{-8}$  or  $10^{-9}$ .

In order to obtain a good estimate of the density of states  $\Omega(E)$  multiple measurements with different sequences of random numbers should be performed and an average of these independent values of  $\Omega(E)$  should be used for the calculation of partition function  $Z(\beta)$  and other relevant thermodynamic quantities. Apart from multiple measurements one may try to minimize correlations between adjacent moves i.e. by increasing the separation between successive records in the histogram  $h(E)$  to obtain a more accurate estimate of  $\Omega(E)$ . In the original Wang Landau algorithm one record is inserted into  $h(E)$  for every trial spin-flip.

Any successful MC simulation algorithm must satisfy the conditions of ergodicity and detailed balance. By ergodicity we mean that starting from any microstate it should be possible to visit any other microstate if the run is long enough. Any algorithm based on single spin moves, like the Metropolis algorithm, satisfies the ergodicity condition. A single spin move may not take the spin system directly to any state starting from a given state but there always is a path connecting two states via a sequence of single step moves.

The condition of detailed balance means that the probabilities of moving into and moving out from a microstate are equal. The random walk process in the WL algorithm does not satisfy this condition initially as the density of states  $\Omega(E)$  is rapidly being modified. In the late stage, when only fine adjustment are made to the density of states, i.e. the modification factor is close to 1, detailed balance is very close to being fulfilled. This can be seen very easily as the transition probability  $p(E_1 \rightarrow E_2)$  is inversely proportional to  $\Omega(E_2)$ :

$$\frac{p(E_1 \rightarrow E_2)}{p(E_2 \rightarrow E_1)} = \frac{\Omega(E_1)}{\Omega(E_2)}, \quad (4)$$

$$i.e. \frac{1}{\Omega(E_1)} p(E_1 \rightarrow E_2) = \frac{1}{\Omega(E_2)} p(E_2 \rightarrow E_1) \quad (5)$$

hence,

$$\sum_{E_1} \frac{1}{\Omega(E_1)} p(E_1 \rightarrow E_2) = \sum_{E_1} \frac{1}{\Omega(E_2)} p(E_2 \rightarrow E_1) \quad (6)$$

The quantity on the L.H.S. is the product of probability of occurrence of the state  $E_1$  and the probability of the transition from  $E_1 \rightarrow E_2$  and is summed over  $E_1$ . Hence it represents the probability of moving into the state  $E_2$ . Similarly the quantity on the R.H.S. represents the same for the reverse transition and the Equation 5 expresses the condition of detailed balance.

### 3. Wang-Landau algorithm for the one-dimensional Lebwohl-Lasher model

#### 3.1 The one-dimensional Lebwohl-Lasher model and the exact results

The Lebwohl-Lasher model is the lattice version of the Maier-Saupe model (Priestly et al., 1976) which describes a nematic liquid crystal in the mean field approximation. This is a system of a one-dimensional array of three-dimensional spins ( $d = 1, n = 3$ , where  $d$  is the system dimensionality and  $n$  is the spin dimensionality) interacting with nearest neighbours via a potential  $-P_2(\cos\theta_{ij})$ , where  $P_2$  is the second Legendre polynomial and  $\theta_{ij}$  is the angle between the nearest neighbour spins  $i$  and  $j$  (the coupling constant in the interaction has been set to unity).

The one-dimensional LL model ( $d = 1, n = 3$ ), has an exact solution and we decided to choose this simple model to apply and test the performance of the WL algorithm for simulation of joint density of states so that a comparison can be made with the exact results available.

The Hamiltonian of the Lebwohl-Lasher model is given by

$$H = - \sum_{\langle i,j \rangle} P_2(\cos \theta_{ij}) \quad (7)$$

where  $P_2$  is the second order Legendre polynomial and  $\theta_{ij}$  is the angle between the nearest neighbour spins  $i$  and  $j$ . The spins are three dimensional and headless, *i.e.* the system has the  $O(3)$  as well as the local  $Z_2$  symmetry characteristic of a nematic liquid crystal. A vector order parameter is inadequate for the system and a traceless second rank tensor  $\underline{Q}$ , as defined below, is used to describe the orientational order of the system (Chaikin & Lubensky, 1995). One uses,

$$Q_{ij} = \frac{1}{N} \sum_{t=1}^N \left( n_i^t n_j^t - \frac{1}{3} \delta_{ij} \right) \quad (8)$$

where  $n_i^t$  is the  $i$ -th component of the unit vector  $\hat{n}$ , which points along the spin at the site  $t$ .  $N$  is the number of particles in the system. In the ordered state  $\langle Q \rangle$  is non-zero. In a coordinate system with the  $Z$ -axis pointing along the direction of molecular alignment (director) the matrix  $\langle Q \rangle$  is diagonal and for a uniaxial system,

$$\langle \underline{Q} \rangle = S \begin{pmatrix} -1/3 & 0 & 0 \\ 0 & -1/3 & 0 \\ 0 & 0 & 2/3 \end{pmatrix} \quad (9)$$

where,

$$S = \frac{1}{2} \langle (3 \cos^2 \theta^t - 1) \rangle = \langle P_2(\cos \theta^t) \rangle \quad (10)$$

where  $\theta^t$  is the angle between a spin and the director. MC simulations demonstrate that a three dimensional Lebwohl-Lasher model ( $d=3, n=3$ ) exhibits a weakly first order transition, characteristic of a nematic-isotropic transition which is available from the Maier-Saupe model of a nematic in the mean field approximation. On the other hand, for lattice dimensionality  $d=2$  and 1, no true long range order is expected since Mermin-Wagner theorem (Mermin & Wagner, 1966) predicts a fluctuation destruction of long range order. The  $d=2$  LL model has been investigated by a number of authors (Kunz & Zumbach, 1992; Mondal & Roy, 2003) and the system shows a behaviour qualitatively similar to the two dimensional XY model. A quasi-long range order has been observed in this system and this is believed to be related to the existence of topological defects in the system (Dutta & Roy, 2004).

The one-dimensional Lebwohl-Lasher model has been simulated by (Chiccoli et al., 1988) and has also been solved exactly (Vuillermot & Romerio, 1973; 1975). The system is known to be disordered at all finite temperatures and critical behaviour is expected only at  $T=0$ , which resembles an one-dimensional Ising model or the one dimensional Heisenberg model. The second rank spin-spin correlation function  $\rho(r)$  is defined as

$$\rho(r) = \langle P_2(\cos \theta(r)) \rangle \quad (11)$$

where  $\theta(r)$  is the angle between two spins,  $r$  lattice spacings apart. In the thermodynamic limit one would expect both  $S$  and  $\rho(r)$  for  $r \rightarrow \infty$  to vanish whereas in finite systems because of finite size effects both quantities may appear to have small but finite values.

Vuillermot and Romerio (Vuillermot & Romerio, 1973; 1975) presented an exact solution of the planar ( $n=2$ ) and spatial ( $n=3$ ) versions of the Lebwohl-Lasher model in one dimension

( $d=1$ ) for a nematic liquid crystal, without periodic boundary conditions. They also calculated the two-molecule correlation functions and have shown that these models do not exhibit any finite temperature order-disorder phase transition.

The partition function  $Z_N(\tilde{K})$  for the  $N$ -particle system is given by

$$Z_N(\tilde{K}) = \tilde{K}^{-N/2} \exp\left[\frac{2}{3}N\tilde{K}\right] D^N(\tilde{K}^{1/2}) \quad (12)$$

where  $\tilde{K}=3/2T$  and  $D$  is the Dawson function (Abramowitz & Stegun, 1970).

$$D(x) = \exp[-x^2] \int_0^x du \exp[u^2]$$

The dimensionless internal energy  $u_N(\tilde{K})$ , the entropy  $S_N(\tilde{K})$  and the specific heat  $C_N(\tilde{K})$  are given by

$$\frac{2U_N(\tilde{K})}{N} = 1 + \frac{3\tilde{K}^{-1}}{2} - \frac{3}{2}\tilde{K}^{-1/2}D^{-1}(\tilde{K}^{1/2}) \quad (13)$$

$$\frac{S_N(\tilde{K})}{N} = \frac{1}{2} + \tilde{K} - \frac{1}{2}\tilde{K}^{1/2}D^{-1}(\tilde{K}^{1/2}) + \ln \left[ \tilde{K}^{-1/2}D(\tilde{K}^{1/2}) \right] \quad (14)$$

and

$$\frac{2C_N(\tilde{K})}{N} = 1 - \tilde{K}^{3/2} \left[ \frac{\tilde{K}^{-1}}{2} - 1 \right] D^{-1}(\tilde{K}^{1/2}) - \frac{1}{2}\tilde{K}D^{-2}(\tilde{K}^{1/2}). \quad (15)$$

The correlation function is given by

$$\rho_N(r) = \left[ \frac{3}{4}\tilde{K}^{-1/2}D^{-1}(\tilde{K}^{1/2}) - \frac{3}{4}\tilde{K}^{-1} - \frac{1}{2} \right]^r \quad (16)$$

### 3.2 Computational details – how to handle a continuous model

In the model we have investigated, spins can take up any orientation in the three dimensional space and the orientation of each spin is stored in terms of the direction cosines ( $l_1, l_2, l_3$ ). The starting configuration has always been chosen as a random one and to generate a new microstate, a randomly selected spin is chosen and each direction cosine is updated as  $l_i \rightarrow l_i + p * r_i$  (for  $i=1,2,3$ ) where  $p$  is a parameter to be chosen according to some criterion and  $r_i$  is a random number between -1 to +1. To preserve the unit magnitude of the spins, ( $l_1, l_2, l_3$ ) is always normalized.

The energy of the system in the LL model is a continuous variable and in one dimension ( $d = 1$ ) it can have any value between  $-L$  to  $L/2$ . To have a discretization scheme for the implementation of the WL algorithm and for an one dimensional random walk in the energy space, we have chosen an energy range from  $(-L$  to  $0)$  and have divided this energy range into  $M$  bins each having a width  $d_e$ .

As already mentioned in the previous section we use  $g(E_i) = \ln \Omega(E_i)$ ,  $\Omega(E_i)$  being the number of micro-states corresponding to the  $i$ -th bin for which the mid-point has the value  $E_i$ . Initially we set all  $g(E_i)$  ( $i=1,M$ ) to zero and the logarithm of the modification factor  $\ln f$  is taken as 1. Whenever a new microstate is generated by rotating a spin, the new system-energy and hence, the macrostate  $j$  is determined. Whether the move is accepted or not is decided according to the WL prescription (Wang & Landau, 2001) for the probability

$$p_{i \rightarrow j} = \min \left( \frac{\Omega(E_i)}{\Omega(E_j)}, 1 \right). \quad (17)$$

If the state  $j$  is accepted, we make  $g(E_j) = g(E_j) + \ln f$  and  $h(E_j) = h(E_j) + 1$ , where  $h(E_j)$  is the histogram count. Otherwise we make  $g(E_i) = g(E_i) + \ln f$  and  $h(E_i) = h(E_i) + 1$ . This procedure is repeated for  $10^4$  MC sweeps (where one MC sweeps consists of  $L$  attempted moves) and the flatness of the histogram is checked and the cycle is repeated till 90% flatness in the histogram is reached. This completes one iteration, following which we reduce the logarithm of the modification factor  $\ln f \rightarrow \ln f/2$ , reset the histogram, and the whole procedure is repeated. For each lattice size we have continued with the iterations till  $\ln f$  gets reduced to  $10^{-9}$ . We have also calculated the quantity  $S$  (Equation 10) which gives us the magnitude of the order parameter obtained from the largest eigenvalue of the ordering matrix defined in Equation 9 and a two-dimensional random walk was performed in the  $(E-S)$  space for this purpose. The flatness check of the two-dimensional histogram thus generated needs to be reconsidered. For a given value of the energy of the system, the order parameter,  $S$  has a distribution over a certain range of values. The whole range of  $S$  is 0 to 1 and in order to perform the two dimensional random walk in the energy-order parameter  $(E-S)$  space we divide the two dimensional space into  $M \times N$  bins. We represent by  $d_\phi$  the bin-widths involving the parameter other than energy in the two-dimensional walk. Each microstate will now correspond to a macrostate labeled by the indices  $i$  and  $j$  and the acceptance probability given by Equation 17 is now modified to

$$p_{ij \rightarrow kl} = \min \left( \frac{\Omega(E_i, S_j)}{\Omega(E_k, S_l)}, 1 \right) \quad (18)$$

along with an appropriate modification of the procedure described after Equation 17 for the two-dimensional random walk. Here, for instance,  $\Omega(E_i, S_j)$  is the density of states for the  $i$ -th energy and  $j$ -th order parameter bin. A two-dimensional random walk in the  $(E-S)$  space is a lot more expensive in terms of the CPU time than an one-dimensional walk in the energy space alone. The problem is particularly severe in a system with continuous energy and becomes worse as the lattice size increases. However, this ensures a much more uniform sampling of the order parameter bins that correspond to a particular energy bin and this improves the overall statistics of the work. It may be pointed out that it is impossible to arrive at a flat histogram in the  $(E-S)$  space if one attempts to visit the entire energy and order parameter ranges accessible to the system. For an one-dimensional walk one normally faces a problem in that, it takes a relatively long time to visit the lowest energy levels and this increases with the increase in system size. For a two-dimensional walk the possibility of uniformly visiting the entire rectangular  $(E-S)$  space is unphysical and one must have a prior knowledge of the range of  $S$ -bins which are likely to be visited while the system energy has a given value  $E_i$ . Our method of simulating the two dimensional random walk has resemblance to the work of Troster and Dellago (Troster & Dellago, 2005), who have applied the WL algorithm to evaluate multidimensional integrals of sharply peaked functions. Our modified approach is elaborated in the following paragraph. We have first mapped the  $(E-S)$  space which costed us  $35 \times 10^6$  sweeps (to be called the pre-production run). The idea is to determine the minimum ( $S_{min}^i$ ) and maximum ( $S_{max}^i$ ) values of the  $S$ -bins which are visited while the system energy is  $E_i$  for  $i=1, M$ . We observe that there are always some  $S$ -bins within the range ( $S_{min}^i, S_{max}^i$ ) for each  $E_i$ , where either no sampling or very low sampling takes place during the pre-production

run. We therefore checked the histograms of the  $(E_i, S_j)$  bins in the mapped region of the two-dimensional space and those which attain a 90% flatness (i.e.  $x = 0.9$ ) during the pre-production run are marked with '1' while other bins are marked '0'. This may be clarified as follows. We calculate the average histogram value for those bins which have been visited at least once, thus discarding the bins which are not visited at all. The flatness test (which needs each of the visited bins to have a histogram count at least equal to 90% of the average histogram) is then applied only to those bins and these are labeled with '1'. In the 'production run' part of the simulation we check the flatness of only those bins which were marked '1' ignoring what is happening to the others. There is however, always a possibility, since the 'production run' generates many more microstates than that in the 'pre-production run', that larger areas in the  $(E-S)$  space would get included in the initial 'visit-map' or those bins, once marked '0', would subsequently qualify for the label '1'. But it is impossible to improve upon the accuracy of the work indefinitely and we decided to stick to the map we obtained during a reasonable amount of the 'pre-production run', ignoring what is happening to the discarded bins. In addition to the two-dimensional random walk in the  $(E-S)$  space we have also performed a number of other two-dimensional random walks. These involve the  $(E-\rho(r))$  space where  $\rho(r)$  is the correlation function defined in Equation 11. Random walks in  $E-\rho(r)$  space are performed only for the  $L=160$  lattice, for  $r$  ranging from 2 to 40 and the ensemble averages of  $\rho(r)$  were evaluated for different temperatures using Equation 3.

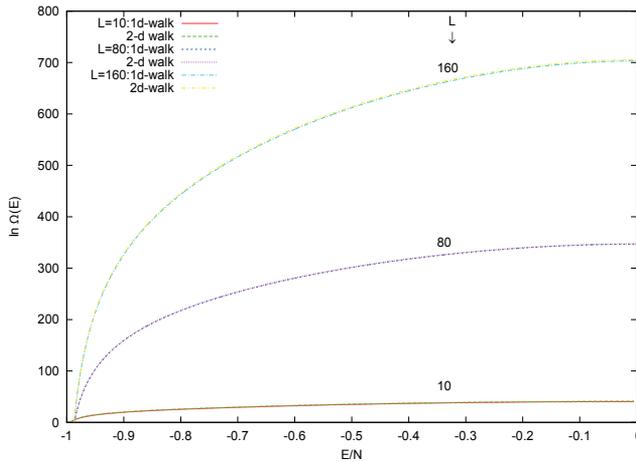


Fig. 1. Logarithm of the density of states,  $\ln \Omega(E)$ , for the 1-d Lebwohl-Lasher model for  $L=10, 80$  and  $160$  obtained from 1-d and 2-d walks. In the resolution of the figure the data for 1-d and 2-d walks overlap.

### 3.3 How to choose different parameters for a continuous model

MC simulations using the WL algorithm in linear spin chains of length  $L$  where  $L=10, 20, 40, 80$  and  $160$  have been performed. All the results we present are results of a single simulation for each lattice size and we did not perform averaging of results over multiple simulations although this surely is expected to reduce the errors. In a simulation involving a continuous model one is confronted with the proper choice of the values of two parameters,  $p$  and  $d_e$ . The former determines the amplitude of the random rotation of a spin and the latter is the energy bin width. In the case of a two dimensional random walk, another parameter  $d_\phi$ ,

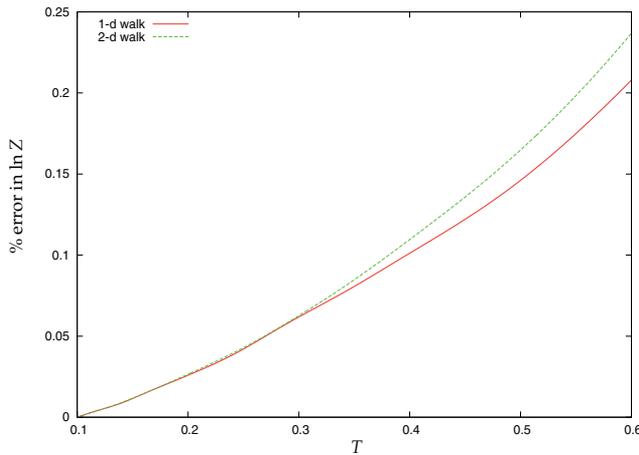


Fig. 2. Percentage error in the logarithm of partition function,  $\ln Z$ , for 1-d and 2-d walks, plotted against temperature  $T$ , for the 1-d Lebwohl-Lasher model for  $L=80$ . The errors are in comparison to the exact results.

which represents the width of the order parameter or the correlation function bin is also to be chosen. We have set  $d_e=0.1$  for all the work reported in this paper.  $d_\phi$  was taken to be 0.01 for both order parameter and correlation function. The parameter  $p$  was always 0.1, except for the two-dimensional walk involving the order parameter, where it was taken to be 0.2. For larger values of  $p$ , the CPU time is less, but the results of the simulation (like the position and height of the specific heat curve, for instance) tend to depend strongly on  $p$ . For the values of  $p$  in the neighbourhood of 0.1, the results depend very weakly on  $p$ . This is presumably due to the fact that, a small change in the orientation of a spin (one at a time) results in a systematic and uniform sampling in the phase space, but as a result of greater correlation of the successive configurations generated and the consequent slow movement of the representative point in phase space, the computer time involved is greater. For the two-dimensional walk involving the order parameter, since a lot of CPU time is necessary, we have chosen  $p=0.2$ , to reduce the time involved. The bin widths for energy or other variables were so chosen that for about 50% of the configurations generated by the spin rotation procedure, new bins are visited. This procedure was found to be optimum, as an attempt to visit new bins more frequently, would result in missing a vast majority of the microstates which correspond to each bin. A small value of  $p$ , the rotation amplitude, is justifiable from the same point of view. A relatively large value of  $p$  results in a poor sampling of the infinite number of closely spaced microstates contained in each bin and leads to poor results which tend to depend on the value of  $p$ , and consequently not in agreement with the exact results.

### 3.4 Results for the one-dimensional LL model: Comparison with exact results

In Fig. 1 we have plotted the quantity  $g(E) = \ln \Omega(E)$  as a function of the energy per particle for  $L=10, 80$  and  $160$  and the results obtained from one and two dimensional walks (in  $E$ - $S$  space) have been compared. The system energy was always considered up to  $E=0$ . The lower limit of the energy for  $L=80$  was  $-79$  and for  $L=160$ , it was  $-158$ , where the corresponding ground state energies are  $-80$  and  $-160$ . Thus, the visited energy range goes to a sufficiently low value to cover the entire range of interest but the small cut near the ground state was necessary, as it takes a huge time to sample these states for a relatively large lattice. The

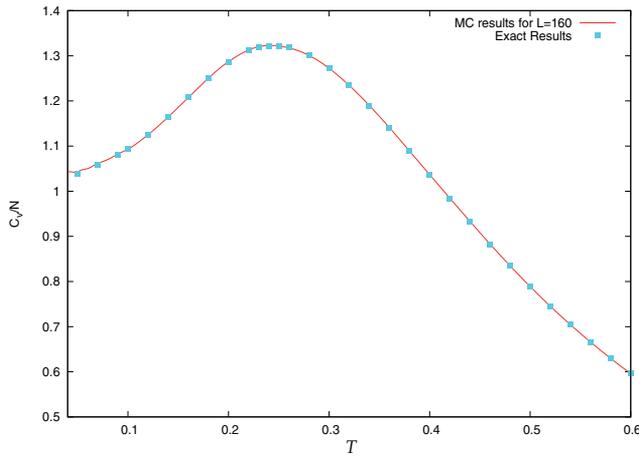


Fig. 3. The specific heat per particle, obtained as a fluctuation quantity, plotted against temperature, for  $L=160$  and compared with the exact result.

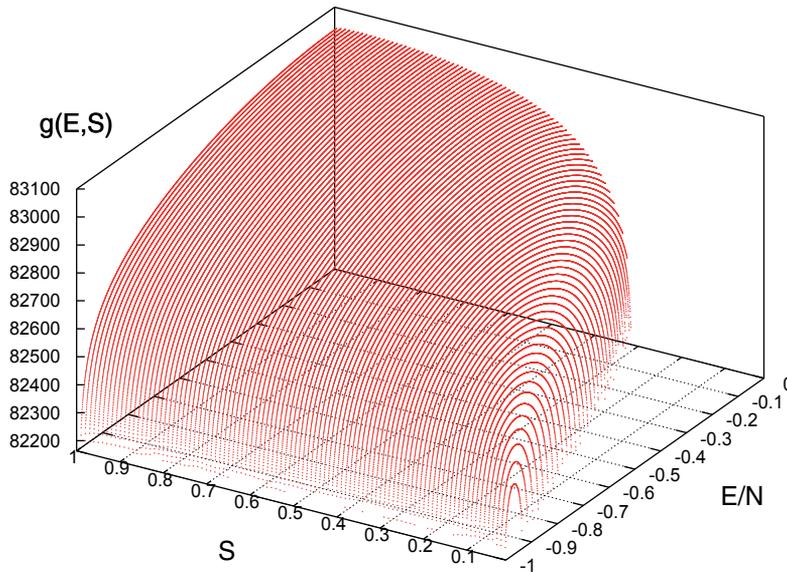


Fig. 4. Three dimensional  $g(E, S)$  surface plotted against energy and the order parameter for  $L=160$ .

partition function,  $Z$  was calculated from a knowledge of the density of states, and the percentage error in  $\ln Z$ , in comparison with the exact results, has been shown in Fig. 2 for both 1-d and 2-d walk for the  $L=80$  lattice. The error in  $\ln Z$  slowly increases with temperature and, at the highest temperature we have investigated, ( $T=0.6$ ), it is about 0.2%; the error in  $\ln Z$  available from 2-d walk being marginally higher.

The specific heat per particle has been plotted against temperature in Fig. 3 for  $L=160$  and compared with the exact results. The surface plot of joint density of states of energy and order parameter for one-dimensional LL model ( $L=160$ ) is shown in Fig. 4. In Fig. 5 we have

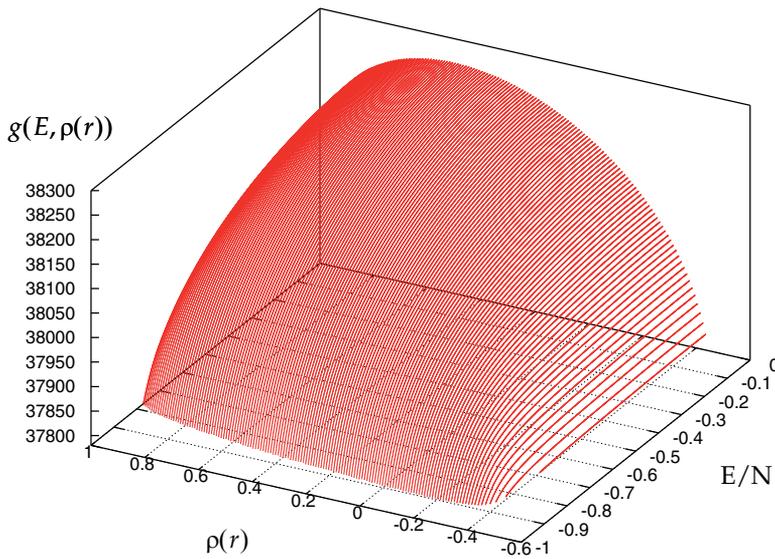


Fig. 5. Three dimensional  $g(E, \rho(r))$  surface plotted against energy per particle and the correlation function for  $r=5$  for the one-dimensional LL model of system size 160.

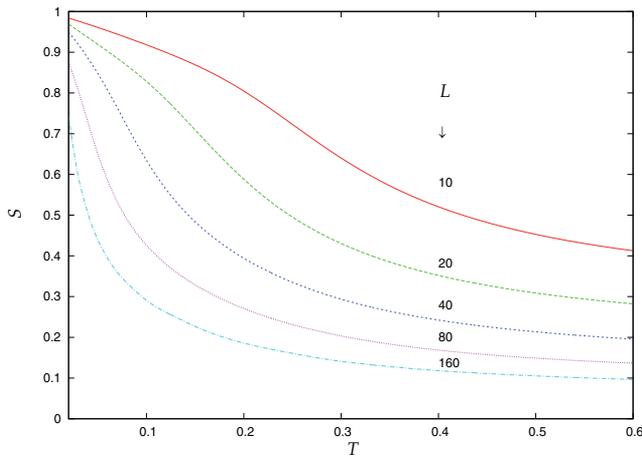


Fig. 6. The temperature variation of orientational order parameter for different lattice sizes for the LL model obtained from JDOS.

depicted the JDOS,  $\ln \Omega[E, \rho(r)]$  for  $r=5$ . The scalar order parameter  $S$ , defined in Equation 10 has been plotted in Fig. 6 against temperature for all lattice sizes. It may be recalled that the system is disordered at all finite temperatures and one would expect  $S=0$  for all values of  $T$ . For a given  $T$  (including  $T=0$ ),  $S$  rapidly falls off with increase in system size, and in the thermodynamic limit will disappear altogether.

The correlation function  $\rho(r)$  has been compared for the  $L=160$  lattice using the two-dimensional random walk. We have performed simulations for 12 values of  $r$ , ranging from 1 to 40, and these have been plotted against temperature in Fig. 7 where comparison

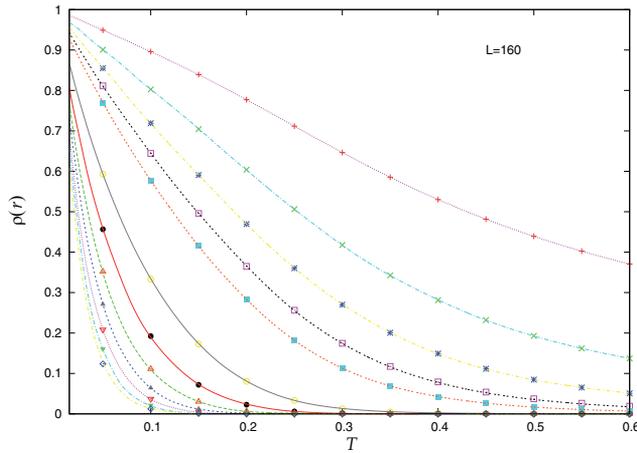


Fig. 7. Variation of correlation function  $\rho(r)$  with temperature  $T$  for lattice size  $L=160$ . The points represented by different symbols are from the exact results. The curves are the results we obtained from the joint density of states. The values of  $r$  taken are 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35 and 40. The topmost curve is for  $r=1$  and the lower curves are for other values of  $r$  given in the sequence above and in ascending order of  $r$ .

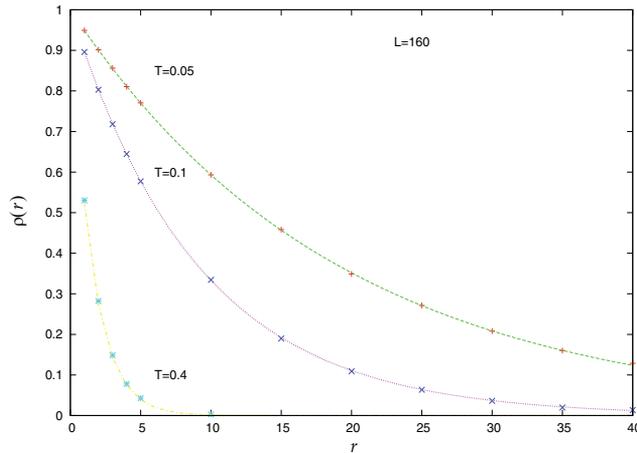


Fig. 8. Correlation function as a function of  $r$  at different temperatures for the LL model obtained using two-dimensional random walk.

has also been made with the exact results. It may be noted that, for each value of  $r$ , we had to run one simulation and the joint density of states were determined separately in each case. The same data has also been shown in Fig. 8 where for three temperatures  $\rho(r)$  has been plotted against  $r$ . As one would expect in a disordered system, the correlation dies off quickly with increase in  $r$ . For  $r \rightarrow \infty$ , the spins are uncorrelated and  $\rho(r)$  should approach  $S^2$ . However, verification of this result from our simulation data will not make much sense in a disordered system. The CPU time necessary for the  $L=160$  lattice for one-dimensional walk is 10.6 hours and the two-dimensional walk involving correlation function is 70.5 hours. For the two-dimensional  $(E,S)$  walk the CPU time is 170 hours. The program was vectorized

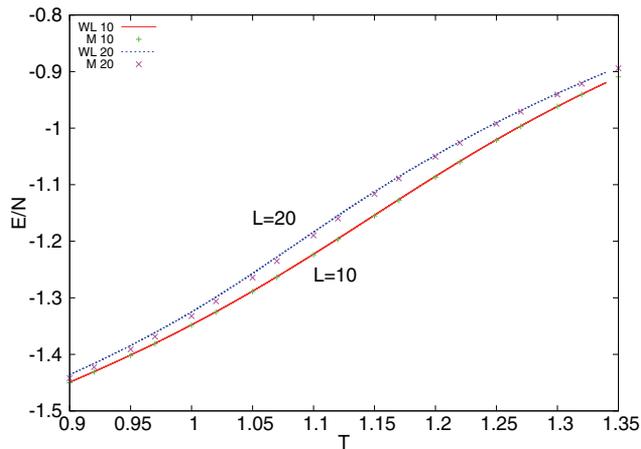


Fig. 9. The average energy per particle for the XY-model plotted against temperature for  $10 \times 10$  and  $20 \times 20$  lattice sizes, obtained using two MC methods namely, WL and M (Metropolis). All results are obtained by taking averages over 20 independent simulations. One-dimensional random walk has been used.

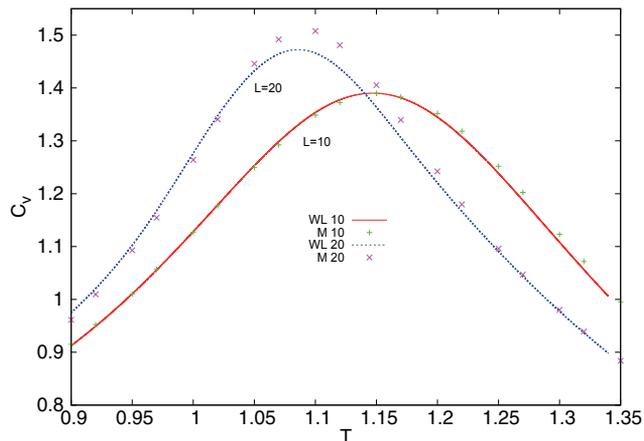


Fig. 10. The specific heat per particle for the XY-model plotted against temperature for  $10 \times 10$  and  $20 \times 20$  lattice sizes, obtained using two MC methods namely, WL and M (Metropolis). All results are obtained by taking averages over 20 independent simulations and one-dimensional random walk with 80% flatness of histogram was used.

(i.e. the 'do' loops parallelized) between two 3.0 GHz Xeon processors in a X226 IBM Server, automatically by the Intel Fortran Compiler, we used.

#### 4. Wang Landau algorithm for two-dimensional XY-model

An XY model is a continuous spin model in which the spins on the lattice are confined to rotate in a plane. Unlike Ising and Potts models which have discrete energy values the XY model has a continuous range of energy values. In this section we consider a two-dimensional square lattice at every site of which is a spin confined to the XY-plane ( $d = 2, n = 2$ ). Note that a

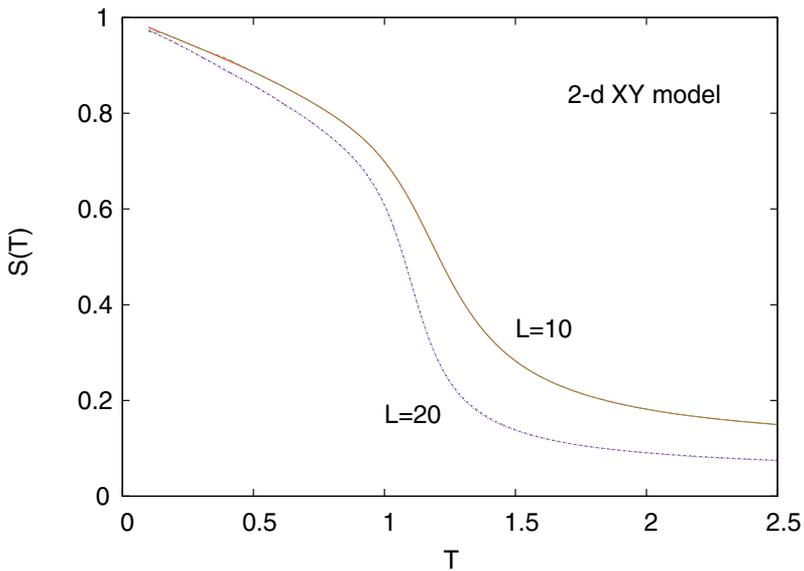


Fig. 11. The orientational order parameter for the XY-model is shown against temperature  $T$  for  $10 \times 10$  and  $20 \times 20$  lattice sizes. The results have been obtained from two-dimensional random walk.

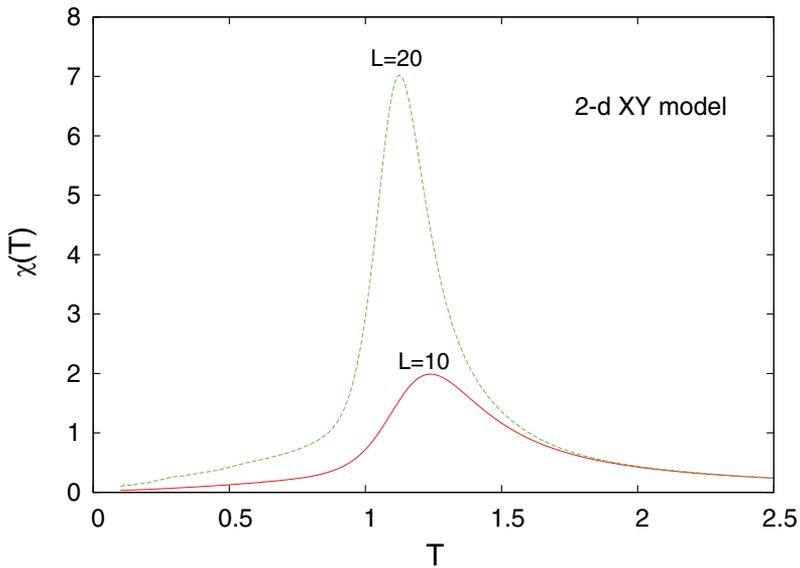


Fig. 12. The susceptibility for the XY-model is shown against temperature  $T$  for  $10 \times 10$  and  $20 \times 20$  lattice sizes.

lattice containing planar spins could have higher dimensions too, for example  $d = 3, n = 2$  is also a possible choice.

#### 4.1 The two-dimensional XY model

In this model planar spins placed at the sites of a planar square lattice interact with nearest neighbours via a potential,

$$V(\theta_{ij}) = -\cos \theta_{ij} \quad (19)$$

where  $\theta_{ij}$  is the angle between the nearest neighbour spins  $i$  and  $j$ . The XY-model is known to exhibit a quasi-long-range-order disorder transition which is mediated by unbinding of topological defects. It has also been the subject of extensive MC simulation over last few decades and some of the recent results may be found in (Bhar & Roy, 2009; Maucourt & Grepel, 1997; Olsson, 1995; Palma et al., 2002).

In this model, the orientational order parameter is defined as follows: let  $\mathbf{n}$  be the unit vector (called the director) in the direction of maximum order prevailing in the system and  $\mathbf{s}$  be the spin vector of unit magnitude, then the average order parameter  $S$  is given by

$$S(T) = \langle \mathbf{n} \cdot \mathbf{s} \rangle = \langle \cos \alpha \rangle \quad (20)$$

where  $\alpha$  is the angle between the director and the spin. The order parameter susceptibility per spin is calculated from the fluctuations in the order parameter using the relation:

$$\chi(T) = \beta N \left( \langle \phi^2 \rangle - \langle \phi \rangle^2 \right) \quad (21)$$

where  $\phi$  is the order parameter for a given configuration and  $N$  is the total number of spins of the system.

#### 4.2 Computational details and results for the two-dimensional XY model

In this model simulations were carried out for lattice sizes  $10 \times 10$  and  $20 \times 20$  and the corresponding minimum of the energy range were chosen to be 3 for each. Histogram flatness was restricted to 80% (i.e.  $x = 0.8$ ) in WL simulation. For the two-dimensional XY model of linear dimension  $L$  the system energy lies between  $-2L^2$  and  $2L^2$ . In order to apply the WL algorithm we have restricted the random walk in the energy space from  $-2L^2$  to 0 (actually a small energy band near the ground state was also excluded to avoid trapping of the random walker in these low energy states as these are scarcely visited during the simulation). The iterations are continued till logarithm of the modification factor  $\ln f$  gets reduced to  $\ln f_{final} = 10^{-8}$  for the  $10 \times 10$  lattice and to  $\ln f_{final} = 10^{-6}$  for the lattice size  $20 \times 20$ . It is not possible to compare the results of our simulation in the XY-model for the DOS or partition function, with the exact results for this model, as we have done for the 1-d LL model, because such exact results for this model are not available. We therefore present here a comparison of WL results with those obtained from simulation using the conventional Metropolis algorithm, which is known to work satisfactorily for this model (Palma et al., 2002). To increase the reliability of our results, we have performed 20 independent simulations for each of the Metropolis, and WL methods and have averaged over the respective results. In the Metropolis simulation the averaging was done directly over the observables while in the other case, the DOS was averaged before it was used to obtain the observables.

Regarding the choice of energy bin-width  $d_e$  and the parameter  $p$ , the considerations as were applied to the 1-d LL model, were taken into account. Unless otherwise stated, we have used  $d_e = 0.1$  and  $p = 0.1$  in this model too. This led us to work with a large number of bins, namely 8000, for the  $20 \times 20$  lattice.

In Figs. 9 and 10 the variations of the average energy,  $E$  and the specific heat,  $C_v$  with temperature are shown respectively for the  $10 \times 10$  and  $20 \times 20$  lattices. These are obtained

with the WL and Metropolis algorithms. This diagram and all the following diagrams, depicting the results of the XY-model, represent the results averaged over 20 independent simulations for the algorithm used. The order parameter  $S$  for two-dimensional XY model of linear lattice sizes 10 and 20 are plotted with temperature in Fig. 11. The system is known to possess no true long range order and a quasi-long range order-disorder transition takes place due to the unbinding of topological defects. Susceptibility of two-dimensional XY model is also plotted as a function of temperature for two lattice sizes in Fig. 12.

## 5. Conclusion

We end this chapter with discussions on the accuracy of the estimated density of states and alternative convergence methods. In course of the random walk in a Wang Landau simulation, the fluctuation of the energy histogram, for a given modification factor  $f$ , initially grows with the number of Monte Carlo sweeps and then saturates to a certain value. Because of the nature of this algorithm the value of the histogram fluctuation determines the error which is generated in the resulting density of states. Zhou and Bhatt carried out a mathematical analysis of the Wang Landau algorithm and proved the convergence of the iterative procedure. They have shown that the error in the density of states, for a given  $f$ , is of the order of  $\sqrt{\ln f}$ . This finding has been tested by Lee *et al.* (Lee et al., 2006). who performed extensive numerical tests in two two-dimensional discrete models, namely, the ferromagnetic Ising model and the fully frustrated Ising model. They have shown that the fluctuation in the histogram increases during an initial accumulation stage and then saturates to a value which is inversely proportional to  $\sqrt{\ln f}$  and they were of the view that this feature is generic to the Wang Landau algorithm. The resulting error in the density of states was found to be of the order of  $\sqrt{\ln f}$ , which is in agreement with the prediction of Zhou and Bhatt.

Convergence methods which are alternatives to the requirement of the histogram flatness, for deciding where to stop an iteration for a given modification factor, have been proposed in the work described in the references (Zhou & Bhatt, 2005), (Lee et al., 2006). According to Zhou and Bhatt an iteration may be stopped when the minimum number of visits to each macrostate is  $1/\sqrt{\ln f}$ . On the other hand, Lee *et al.* proposed that an iteration can be stopped when the number of Monte Carlo sweeps, for a given value of  $f$ , is such that the saturation of the histogram fluctuation has been reached, since continuing the simulation for this particular modification factor is unlikely to reduce the error in the density of states any further. Sinha and Roy (Sinha & Roy, 2009) have investigated the growth of histogram fluctuations in two continuous lattice spin models. In these models where the spins reside at the sites of a two-dimensional square lattice, the interaction between nearest neighbours is given by

$$V(\theta_{ij}) = 2 \left[ 1 - \left( \cos^2 \frac{\theta_{ij}}{2} \right)^{q^2} \right]$$

where  $q^2$  is a positive number and  $\theta_{ij}$  is the angle between the nearest neighbour spins  $i$  and  $j$ . For  $q^2 = 1$ , the model is simply the conventional XY model (except for an additive constant in the potential). This model has also been studied recently by Bhar and Roy (Bhar & Roy, 2009) using WL and Wang-Landau-Transition-Matrix Monte Carlo algorithm. For larger  $q^2$  (say,  $q^2 = 50$ ) the potential well has a sharp minimum and the system possesses a strongly first order phase transition. Sinha and Roy have observed that in these continuous models too, the fluctuations in the energy histogram after an initial increase saturates to a value proportional to  $1/\sqrt{\ln f}$ . Therefore, it may seem, in agreement with the proposal of Lee et al., that WL

sampling should be carried out only till the saturation value of the histogram sampling is reached. However it was found that this does not work even for systems of moderate size as an energy range near the ground state is not sampled at all. Even the idea of Zhou and Bhatt that  $1/\sqrt{\ln f}$  visits to each bin does not seem to be a practical solution as it involves a huge CPU time even for continuous systems of moderate size. This remains a major problem with the feasibility of applying the WL method to continuous spin systems and problem is more tedious when one is determining the JDOS.

## 6. Acknowledgment

We acknowledge the receipt of a research grant No. 03(1071)/06/EMR – II from the Council of Scientific and Industrial Research, India which help us to procure the IBM X226 servers. One of the authors (NG) acknowledges the award of a teacher fellowship under Faculty Development Programme of the UGC, India.

## 7. Appendix

In this appendix we show the technique for the evaluation of the partition function  $Z$  from the estimated density of states. The direct use of Equation 1 in evaluating partition function may cause overflow or underflow problem in the allowed range of real numbers on modern computers. In order to get around of this problem we may use the following trick. Suppose we have a system with four energy levels:  $E_1, E_2, E_3$  and  $E_4$  and the logarithms of the corresponding density of states are  $g_1, g_2, g_3$  and  $g_4$  the values of which are generated in simulations. Then the partition function of the system at temperature  $T$  is expressed as

$$Z = e^{v_1} + e^{v_2} + e^{v_3} + e^{v_4} \quad (22)$$

where  $v_i = g_i - \beta E_i, \quad i = 1, 2, 3, 4.$

$$\begin{aligned} Z &= e^{v_1} [e^{v_2-v_1} + e^{v_3-v_1} + e^{v_4-v_1}] \\ &= e^{v_1} [1 + e^{v_2-v_1} [1 + e^{v_3-v_2} + e^{v_4-v_2}]] \\ &= e^{v_1} [1 + e^{v_2-v_1} [1 + e^{v_3-v_2} [1 + e^{v_4-v_3}]]] \end{aligned} \quad (23)$$

This can further be written as

$$\begin{aligned} Z &= e^{v_1} \left[ 1 + e^{v_2-v_1} \left[ 1 + e^{[v_3-v_2+\ln[1+e^{v_4-v_3}]]} \right] \right] \\ &= e^{v_1} \left[ 1 + e^{\left[ v_2-v_1+\ln \left[ 1+e^{\left[ v_3-v_2+\ln \left[ 1+e^{v_4-v_3} \right] \right]} \right] \right]} \right] \end{aligned} \quad (24)$$

Therefore

$$\ln Z = v_1 + e^{v_1} \left[ 1 + e^{\left[ v_2-v_1+\ln \left[ 1+e^{\left[ v_3-v_2+\ln \left[ 1+e^{v_4-v_3} \right] \right]} \right] \right]} \right] \quad (25)$$

Since the exponents of the exponential are now written as differences of two terms the problem of overflowing is avoided. Again, if each of the powers of the exponential in the above

equation are greater than a predefined large value say, 500, we can neglect 1. This also simplifies the evaluation of  $Z$ .

## 8. References

- Abramowitz, M. & Stegun, I.(1970). *Handbook of Mathematical Functions*, Dover, ISBN 486-61272-4, New York.
- Bhar, S. & Roy, S.K., (2009). Computer simulation of two continuous spin models using Wang-Landau-Transition- Matrix Monte Carlo algorithm. *Comput. Phys. Comm.* Vol. 180, pp. 699-707.
- Chaikin, P. M. & Lubensky, T. C. (1995). *Principles of Condensed Matter Physics*, Cambridge Univ. Press , ISBN 81-7596-025-6, Cambridge.
- Chiccoli, C.; Pasini, P. & Zannoni, C. (1988). Can Monte Carlo detect the absence of ordering in a model liquid crystal? *Liq. Cryst.*, Vol. 3, pp. 363-368.
- de Oliveira, P. M. C.; Penna, T. J. P. & Herrmann, H. J. (1996). Broad histogram method. *Braz. J. Phys.*, Vol. 26, pp. 677-683.
- Dutta, S. & Roy, S. K. (2004). Phase transitions in two planar lattice models and topological defects: A Monte Carlo study. *Phys. Rev. E.*, Vol. 70, pp. 066125-1-9.
- Edwards, S. F. & Anderson, P. W. (1975). Theory of spin glasses. *J. Phys. F.*, Vol. 5, pp. 965-974.
- Ferrenberg, A. M. & Swendsen, R. H. (1988). New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett.*, Vol. 61, pp. 2635-2638.
- Ferrenberg, A. M. & Swendsen, R. H. (1988). Optimized Monte Carlo data analysis. *Phys. Rev. Lett.*, Vol. 63, pp. 1195-1198.
- Kosterlitz, J. M. & Thouless, D. J. (1973). Ordering, metastability and phase transitions in two-dimensional systems. *J. Phys. C:Solid State Phys.*, Vol. 6, pp. 1181-1203.
- Kosterlitz, J. M. (1974). The critical properties of the two- dimensional xy model. *J. Phys. C:Solid State Phys.*, Vol. 7, pp. 1046-1060.
- Kunz, H. & Zumbach, G. (1992). Topological phase transition in a two-dimensional nematic n-vector model: A numerical study. *Phys. Rev. B.*, Vol. 46, No. 2, pp. 662-673.
- Landau, D. P. & Binder, K. , (2000). *A Guide to Monte Carlo Methods in Statistical Physics*, Cambridge Univ. Press., ISBN 13-978-0-521-84238-9, New York.
- Lebwohl, P.A. & Lasher, G. (1972). Nematic-Liquid-Crystal Order- A Monte Carlo Calculation . *Phys. Rev. A.*, Vol. 6, No. 1, pp. 426-429.
- Lee, H.K.; Okabe, Y. & Landau, D.P., (2006). Convergence and refinement of the Wang-Landau algorithm. *Computer Physics Communications*, Vol. 175, No. 36, pp. 36-40.
- Maucourt, J. & Gempel, D. R. (1997). Phase transitions in the two-dimensional XY model with random phases: A Monte Carlo study. *Phys. Rev. B.*, Vol. 56, pp. 2572-2579.
- Mermin, N.D. & Wagner, H. (1966). Absence of ferromagnetism or antiferromagnetism in one- or two- dimensional isotropic Heisenberg models. *Phys. Rev. Lett.*, Vol. 17, No. 22 pp. 1133-1136.
- Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.;Teller, A. H. & Teller, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.*, Vol. 21, pp. 1087-1092.
- Mondal, E. & Roy, S. K. (2003). Finite size scaling in the planar Lebwohl-Lasher model. *Phys. Lett. A.*, Vol. 312, pp. 397-410.
- Mukhopadhyay, K.; Ghoshal, N. & Roy, S.K., (2008). Monte Carlo simulation of joint density of states in one-dimensional Lebwohl-Lasher model using Wang-Landau algorithm. *Phys. Lett. A*, Vol. 372, pp. 3369-3374.

- Newman, M. E. J. & Barkema, G. T. , (1999). *Monte Carlo Methods in Statistical Physics*, Clarendon Press, ISBN 0-19-851796-3, Oxford.
- Olsson, P. (1995). Monte Carlo analysis of the two-dimensional XY model. II Comparison with the Kosterlitz renormalization-group equations. *Phys. Rev. B.*, Vol. 52, pp. 4526-4535.
- Onsager, L. (1944). Crystal statistics. I. A two-dimensional model with an order-disorder transition. *Phys. Rev.*, Vol. 65, pp. 117-149.
- Palma, G.; Mayer, T. & Labbe, R. (2002). Finite size scaling in the two-dimensional XY model and generalised universality. *Phys. Rev. E.*, Vol. 66, pp. 026108-1-5.
- Priestly, E.B.; Wojtowicz, P.J. & Sheng, P. (1976). *Introduction to Liquid Crystals*, Plenum Press, ISBN 0-306-30858-4, New York.
- Sinha, S. & Roy, S.K., (2009). Performance of Wang-Landau algorithm in continuous spin models and a case study: Modified XY model. *Phys. Lett. A*, Vol. 373, pp. 308-314.
- Swendsen, R. H. & Wang, J. S. (1987). Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.* , Vol. 58, pp. 86-88.
- Troster, A. & Dellago, C. (2005). Wang-Landau sampling with self-adaptive range. *Phys. Rev. E.*, Vol. 71, No. 066705, pp. 1-7.
- Vuillermot, P. A. & Romerio, M. V. (1973). Exact solution of the Maier-Saupe model for a nematic liquid crystal on a one-dimensional lattice. *J. Phys. C: Solid State Physics*, Vol. 6, pp. 2922-2930.
- Vuillermot, P. A. & Romerio, M. V. (1975). Absence of Ordering in a Class of Lattice Systems. *Commun. Math. Phys.*, Vol. 41, pp. 281-288.
- Wang, F. & Landau, D.P. (2001). Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States. *Phys. Rev. Lett.*, Vol. 86, pp. 2050-2053; Wang, F. & Landau, D.P. (2001). Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram. *Phys. Rev. E.*, Vol. 64, No. 056101, pp. 1-16.
- Wolff, U. (1989). Collective Monte Carlo updating for spin systems. *Phys. Rev. Lett.*, Vol. 62, pp. 361-364.
- Wu, F. Y. (1982). The Potts model. *Rev. Mod. Phys.*, Vol. 54, pp. 235-268.
- Zhou, C. & Bhatt, R.N. (2005). Understanding and improving the Wang-Landau algorithm. *Phys. Rev. E.*, Vol. 72, No. 025701, pp. 1-4.

# Characterizing Molecular Rotations using Monte Carlo Simulations

Bart Verberck  
*University of Antwerp*  
*Belgium*

## 1. Introduction

The Monte Carlo (MC) simulation technique is a powerful method for calculating thermodynamic averages of physical quantities of many-body systems. The physical property we focus on in this Chapter is molecular rotational motion. The rotational degrees of freedom in molecular crystals give rise to temperature- and/or pressure-driven transitions between phases with freely rotating, quasi-freely rotating, or orientationally ordered molecules. Orientationally disordered crystals represent a state of matter between the liquid and the purely crystalline state, and can be compared to liquid crystals. In liquid crystals, however, translational order is destroyed and orientational order is preserved, while in molecular crystals translational order persists while molecules are (partially) orientationally disordered. For a review on molecular crystals we refer to Lynden-Bell & Michel (1994). The molecular crystals we envisage can be as simple as solid hydrogen, but as complex as protein crystals. Also, we do not restrict ourselves to crystals containing only one type of molecule, or to three-dimensional (3D) molecular arrangements. An example of a heterogeneous molecular crystal is fullerene-cubane,  $C_{60} \cdot C_8H_8$ , while fullerene molecules like  $C_{60}$  or  $C_{70}$  packed inside a carbon nanotube (CNT) provide an instance of a one-dimensional (1D) molecular chain. MC simulations provide an excellent tool for the computational study of the different phases, and the transitions between them, of molecular crystals. First, molecular crystals typically consist of molecules interacting via van der Waals interactions, which can be relatively easily modeled using phenomenological potential models. Secondly, the main advantage of MC simulations is the possibility to directly change pressure and temperature, and to examine how the crystal's structure (from the point of view of molecular order/disorder) changes accordingly.

While the actual implementation of molecular rotations in (MC) simulations is typically covered in textbooks, e.g. Allen & Tildesley (1987) and Frenkel & Smit (2002), the actual characterization of molecular rotations and orientations has received much less attention. In this Chapter, we present a method to assess molecular rotational motion within molecular crystals based on the concept of orientational mean-squared displacements (OMSDs). The technique provides an efficient way for describing different rotational regimes of individual molecules and of the molecular crystal as a whole. From a computational point of view, the method has the advantage that only a limited number of parameters has to be sampled and stored to obtain the necessary information on molecular motion and ordering.

We consider rigid molecules, so that each molecule has three translational and three rotational degrees of freedom. The context of the present Chapter assumes a fully set-up MC simulation of a system of molecules, in any ensemble, with all the usual ingredients like interaction

potentials, periodic boundary conditions, minimum-image convention, trial moves, etc. included. In the next section, we formally define OMSDs and demonstrate how they can be used to quantify molecular motion. In Sect. 3, we turn to the practical implementation of the OMSD method. We also recall how to perform rotational MC trial moves and provide a memory-efficient way of doing simulation runs. Next (Sect. 4), we present two examples where OMSDs have been used to extract information on molecular orientations. In Sect. 5 we provide a Chapter summary.

## 2. OMSDs: general formulation

### 2.1 Definition

During a MC simulation, a sequence of orientations is generated for every molecule. To fix ideas, let us focus on one type of molecule. The concepts introduced here are easily generalizable to multi-component molecular crystals. As a molecule adopts various orientations, for any point  $\vec{r}_i$  fixed with respect to the molecule, e.g. an atom or a bond (when speaking of a bond and its coordinates, the center of the bond is understood), a set of locations  $\{\vec{r}_i(p), p = 1, \dots, P\}$  is produced. Here,  $i$  labels the considered molecule,  $p$  labels subsequent MC steps and  $P$  is the total number of MC samples. The coordinates  $\vec{r}_i = (x_i, y_i, z_i)$  of this “monitored” point are defined with respect to the local cartesian system of axes  $(o, x, y, z)$ , where the origin  $o$  coincides with the molecule’s center of mass, and the axes  $(x, y, z)$  are fixed and parallel to the global coordinate system’s axes  $(X, Y, Z)$ .

For a freely rotating molecule, the set  $\{\vec{r}_i(p)\}$  eventually (for  $P \rightarrow \infty$ ) covers a sphere with radius  $|\vec{r}_i|$ . If we consider the same monitored point for every molecule,  $|\vec{r}_i| \equiv r$  is independent of the molecular index  $i$ . In fact, it is advisable to use the same monitored point for every molecule, and we will work under this assumption throughout the whole Chapter. If the molecule does not rotate at all,  $\vec{r}_i(p) \equiv \vec{r}_i$  is constant. In Fig. 1, the example of a rotating square is shown, with the middle point of one of the edges as the monitored point.

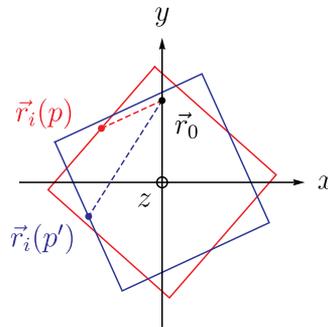


Fig. 1. Two configurations  $p$  and  $p'$  of a rotating square. The middle point of one of the edges is chosen as the monitored point  $\vec{r}_i$ , the fixed point  $\vec{r}_0$  lies on the  $y$ -axis. When the square adopts various orientations, the distance  $u_i = |\vec{r}_i - \vec{r}_0|$  between  $\vec{r}_i$  and  $\vec{r}_0$  (dashed lines) varies accordingly. If the square rotates freely in three dimensions, the monitored point  $\vec{r}_i$  describes a sphere with radius  $|\vec{r}_i|$ .

The idea behind OMSDs is to calculate for every molecule, after each MC step, the square of the distance  $u_i(p)$  between the monitored point  $\vec{r}_i(p)$  and a fixed point  $\vec{r}_0 = (x_0, y_0, z_0)$ :

$$u_i(p)^2 = |\vec{r}_i(p) - \vec{r}_0|^2 = [x_i(p) - x_0]^2 + [y_i(p) - y_0]^2 + [z_i(p) - z_0]^2. \quad (1)$$

The fixed point  $\vec{r}_0$  has to be defined with respect to the local coordinate system  $(o, x, y, z)$  associated with the molecule  $i$  under consideration. We take the same fixed point for every molecule. In Fig. 1 the fixed point  $\vec{r}_0$  is chosen on the  $y$ -axis. The OMSD associated with the chosen fixed point is then obtained by averaging over all MC steps:

$$\langle u_i^2 \rangle = \frac{1}{P} \sum_{p=1}^P u_i(p)^2. \quad (2)$$

It is also useful to average over all molecules in the system:

$$\langle\langle u^2 \rangle\rangle = \frac{1}{N} \sum_{i=1}^N \langle u_i^2 \rangle, \quad (3)$$

with  $N$  the total number of molecules. As we will illustrate in Sect. 4, the overall OMSD  $\langle\langle u^2 \rangle\rangle$  [Eq. (3)] is a good measure for oriental motion in the system when all molecules rotate simultaneously in a similar way, while the molecular OMSDs  $\langle u_i^2 \rangle$  [Eq. (2)] can be used to characterize orientationally ordered phases.

## 2.2 Characterization of orientational regimes using OMSDs

By carefully choosing the monitored and fixed points, it is possible to compare the numerically obtained OMSDs  $\langle u_i^2 \rangle$  with analytically calculated values. As an example, let us choose the point  $(0, 0, z_0)$  as the fixed point, and consider three special cases of molecular motion: (i) free three-dimensional (3D) rotation, (ii) free rotation about the  $z$ -axis and (iii) no rotation (i.e. a fixed orientation).

**Free 3D rotation.** The analytical calculation of  $\langle u_i^2 \rangle$  is most easily done by introducing spherical coordinates:

$$x_i = r_i \sin \theta_i \cos \phi_i, \quad (4a)$$

$$y_i = r_i \sin \theta_i \sin \phi_i, \quad (4b)$$

$$z_i = r_i \cos \theta_i. \quad (4c)$$

Here,  $\vec{r}_i = (x_i, y_i, z_i)$  stands for the monitored point of molecule  $i$ . The squared distance  $u_i^2$  then reads

$$u_i^2 = r_i^2 - 2r_i z_0 \cos \theta_i + z_0^2. \quad (5)$$

To calculate the analytical OMSD, we apply the formula for the 3D orientational average of a quantity  $f \equiv f(\theta_i, \phi_i)$ ,

$$\langle f \rangle_a = \frac{1}{4\pi} \int_0^{2\pi} d\phi_i \int_0^\pi \sin \theta_i d\theta_i f(\theta_i, \phi_i), \quad (6)$$

for  $f = u_i^2$ . The result reads

$$\langle u_i^2 \rangle_a = r_i^2 + z_0^2. \quad (7)$$

Here, we write  $\langle u_i^2 \rangle_a$  for the analytical result to distinguish it from the numerical result  $\langle u_i^2 \rangle$  [Eq. (2)]. The former follows from an integration, the latter from a summation.

**Free rotation about the z-axis.** In the case of free rotation about the z-axis, the average reads

$$\langle f \rangle_a = \frac{1}{2\pi} \int_0^{2\pi} d\phi_i f(\phi_i), \quad (8)$$

which for  $f = u_i^2$  results in

$$\langle u_i^2 \rangle_a = r_i^2 - 2r_i z_0 \cos \theta_i + z_0^2. \quad (9)$$

**Fixed orientation.** If the molecule does not rotate, one simply has

$$\langle f \rangle_a = f, \quad (10)$$

so that

$$\langle u_i^2 \rangle_a = r_i^2 - 2r_i z_0 \cos \theta_i + z_0^2. \quad (11)$$

A characterization of molecular motion can be obtained by comparing the numerical and analytical OMSD values,  $\langle u_i^2 \rangle$  and  $\langle u_i^2 \rangle_a$ , respectively. However, the resulting values do not uniquely define an orientational regime. In the analytical analysis shown above, one finds the same value for free rotation about the z-axis ('free z-rotation') and for no rotation. The remedy for this is to introduce additional fixed points. Repeating the analytical calculation for fixed points  $(x_0, 0, 0)$  and  $(0, y_0, 0)$  shows that with more than one fixed point, a distinction between the three regimes considered above is possible. The resulting analytical OMSD values for  $\vec{r}_0 = (x_0, 0, 0)$ ,  $(0, y_0, 0)$  and  $(0, 0, z_0)$  are summarized in Table 1. Note that the use of only  $(x_0, 0, 0)$  and  $(0, 0, z_0)$  already allows to distinguish between the three proposed rotational regimes, but that the use of only  $(x_0, 0, 0)$  and  $(0, y_0, 0)$  does not. This shows that it is important to decide beforehand which fixed points to use. To avoid ambiguous situations, it is best to implement the calculation of OMSDs based on three fixed points of the type  $(x_0, 0, 0)$ ,  $(0, y_0, 0)$  and  $(0, 0, z_0)$ .

$\vec{r}_0$	$\langle u_i^2 \rangle_a$		
	free 3D rotation	free z-rotation	no rotation
$(x_0, 0, 0)$	$r_i^2 + x_0^2$	$r_i^2 + x_0^2$	$r_i^2 - 2r_i x_0 \sin \theta_i \cos \phi_i + x_0^2$
$(0, y_0, 0)$	$r_i^2 + y_0^2$	$r_i^2 + y_0^2$	$r_i^2 - 2r_i y_0 \sin \theta_i \sin \phi_i + y_0^2$
$(0, 0, z_0)$	$r_i^2 + z_0^2$	$r_i^2 - 2r_i z_0 \cos \theta_i + z_0^2$	$r_i^2 - 2r_i z_0 \cos \theta_i + z_0^2$

Table 1. Analytical OMSD values  $\langle u_i^2 \rangle_a$  for three well-chosen fixed points  $\vec{r}_0$ . Note that for free rotation about the z-axis, the value of  $\theta_i$  is indeed fixed and not averaged out. Likewise for  $\theta_i$  and  $\phi_i$  in the case of no rotation.

### 3. OMSDs: implementation

#### 3.1 Implementation of random rotations in a MC simulation

Whereas performing translational moves of molecules in a MC simulation is straightforward, implementing 3D orientational motion is less trivial. First, molecular orientational jumps in a MC simulation have to be generated in a random way, and secondly, it has to be possible to attribute a controllable amplitude to a rotation. Therefore, the explicit use of Euler rotations is not recommended. Popular modern approaches are based on quaternions — for more details we refer to Allen & Tildesley (1987), Frenkel & Smit (2002) and Vesely (1982). We suggest the following procedure: (i) first select a random unit vector  $\vec{s} \equiv (s_x, s_y, s_z)$  originating from the molecule's center of mass and (ii) then rotate the molecule about  $\vec{s}$  over  $\alpha$ . The angle  $\alpha$  is then

a direct measure for the rotational jumps' amplitude and the interval where its value should be randomly selected from can be adjusted — before definitive data collecting — to yield the desired acceptance rate (typically 50%) for rotational MC trial moves.

The generation of a random unit vector in three dimensions  $\vec{s} = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)$ , in other words randomly choosing a point on a sphere with radius 1, is not achieved by choosing uniformly random values for the polar and azimuthal angles  $\theta$  and  $\phi$  in the intervals  $[0, \pi]$  and  $[0, 2\pi[$ , respectively [Miles (1965)]. (Note that in two dimensions it does suffice to randomly choose  $\psi$  between 0 and  $2\pi$  for  $\vec{s} \equiv (s_x, s_y) = (\cos \psi, \sin \psi)$  to be a random unit vector.) The correct recipe reads

$$s_x = \sqrt{1 - q_1^2} \cos q_2, \quad (12a)$$

$$s_y = \sqrt{1 - q_1^2} \sin q_2, \quad (12b)$$

$$s_z = q_1, \quad (12c)$$

$$q_1 = 2r_1 - 1, \quad (12d)$$

$$q_2 = 2\pi r_2, \quad (12e)$$

where  $r_1$  and  $r_2$  are random numbers uniformly distributed in the interval  $[0, 1[$ . Rotating the point  $\vec{r} \equiv (x, y, z)$  of the molecule about  $\vec{s}$  over  $\alpha$  results in the point

$$\vec{r}' = \vec{r}_{\parallel} + \vec{r}_{\perp} \cos \alpha + (\vec{r}_{\perp} \times \vec{s}) \sin \alpha, \quad (13)$$

where

$$\vec{r}_{\parallel} = (\vec{s} \cdot \vec{r})\vec{s}, \quad (14a)$$

$$\vec{r}_{\perp} = \vec{r} - \vec{r}_{\parallel}. \quad (14b)$$

In the more common form of a matrix multiplication, one has

$$\vec{r}' = R_{\vec{s}}(\alpha)\vec{r} \quad (15)$$

with  $R_{\vec{s}}(\alpha)$  given by

$$R_{\vec{s}}(\alpha) = \begin{pmatrix} s_x^2(1 - \cos \alpha) + \cos \alpha & s_x s_y(1 - \cos \alpha) + s_z \sin \alpha & s_z s_x(1 - \cos \alpha) - s_y \sin \alpha \\ s_x s_y(1 - \cos \alpha) - s_z \sin \alpha & s_y^2(1 - \cos \alpha) + \cos \alpha & s_y s_z(1 - \cos \alpha) + s_x \sin \alpha \\ s_z s_x(1 - \cos \alpha) + s_y \sin \alpha & s_y s_z(1 - \cos \alpha) - s_x \sin \alpha & s_z^2(1 - \cos \alpha) + \cos \alpha \end{pmatrix}. \quad (16)$$

### 3.2 Sampling OMSDs. Resetting the monitored point

The following remarks apply generally to any molecular crystal, possibly with several types of molecules in it, but to fix ideas, we take the example of a system of benzene molecules,  $C_6H_6$ . Let us consider one molecule, labeled  $i$ , and take one of its six C atoms as the monitored point and label that atom  $C_1$ . After starting the MC simulation, it is likely that the point  $C_1$  will drift away from its initial position. At equilibrium, the molecule might rotate three-dimensionally, or rotate about an axis, or adopt a fixed orientation. In any case, it is important to start sampling OMSDs only after equilibrium has set in — which applies in fact for all quantities sampled in a MC simulation — since the path the atom  $C_1$  has been describing before equilibrium is bound to bias the outcome of the value of  $\langle u_1^2 \rangle$ .

A more subtle issue arises from molecular symmetry and the equivalence of points related by symmetry operations. Continuing with the benzene example, we now consider two molecules

$i$  and  $j$ , and monitor their  $C_1$  atoms. Again, after the start of the MC simulation, both atoms are likely to drift away from their initial positions. Note that the initial positions can coincide (in the case of identical molecular orientations at initialisation), but that they do not have to (in the case of random molecular orientations at start-up). If, after equilibrium has set in, the molecules adopt the same fixed orientation, it is probable that atom  $C_1$  of molecule  $i$  and atom  $C_1$  of molecule  $j$  are differently positioned (Fig. 2). This typically occurs upon cooling: at high temperature the molecules rotate, and at low temperature the system is orientationally ordered. In the latter case, different positions of  $C_1$  for molecules  $i$  and  $j$  result in different OMSD values for molecules  $i$  and  $j$ , although their orientations are the same.

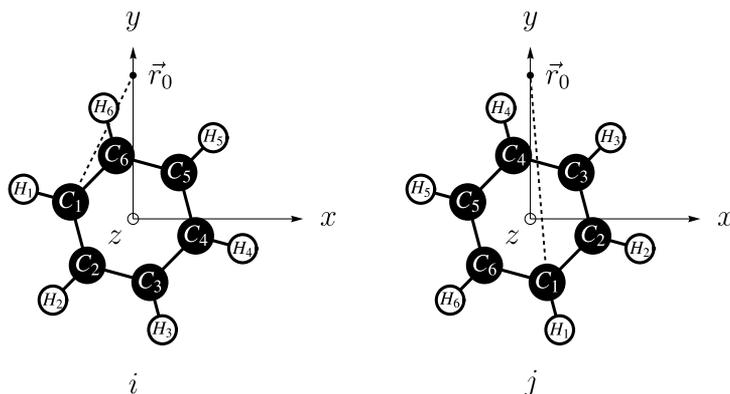


Fig. 2. Benzene molecules  $i$  (left) and  $j$  (right) in equivalent orientations, but with the atoms  $C_k$  and  $H_k$  ( $k = 1, \dots, 6$ ) differently located. With  $C_1$  as the monitored point, the distances  $u_i$  and  $u_j$  (dashed lines) between the monitored point and the fixed point  $\vec{r}_0$  (here chosen on the  $y$ -axis) are different.

A remedy for this is to introduce a fixed reference orientation, which we call standard orientation, and to “reset” the monitored point when starting to sample OMSDs. In the present example, we define the standard orientation as the configuration with all atoms in the  $z$ -plane, and two C atoms lying on the  $y$ -axis (Fig. 3). The C atom with  $y > 0$  is labeled  $C_1$  and chosen to be the **reference monitored point**. When equilibrium has set in, the C atom of the molecule — in its present orientation — closest to the  $C_1$  atom of the molecule in the standard orientation is chosen as the **dynamic monitored point**  $\vec{r}_i$ . In this way, molecular symmetry is accounted for, and different realizations of one molecular configuration result in the same  $u_i$  values — and consequently, in the same OMSD values. Note the importance of the choice of a standard orientation, and the dynamic redefinition of the monitored point.

The foregoing is valid for any choice of monitored point — atoms, bonds, or any other locus. Rather than choosing one particular point (e.g. a C atom), a family of equivalent points (e.g. six C atoms) related by symmetry operations (e.g. six-fold rotations about the  $z$ -axis) is chosen; the point then actually monitored is the point closest to one of the equivalent points (e.g.  $C_1$ ) having a precise location for the molecule in the standard orientation. We point out that the reset procedure is required for every molecule in the system.

### 3.3 Program flow and data management

Having defined OMSDs, discussed how to use them for characterizing rotational regimes, and having made some practical remarks on random rotations and resetting the monitored points, we now turn to a discussion of how to practically include the calculation of OMSDs in

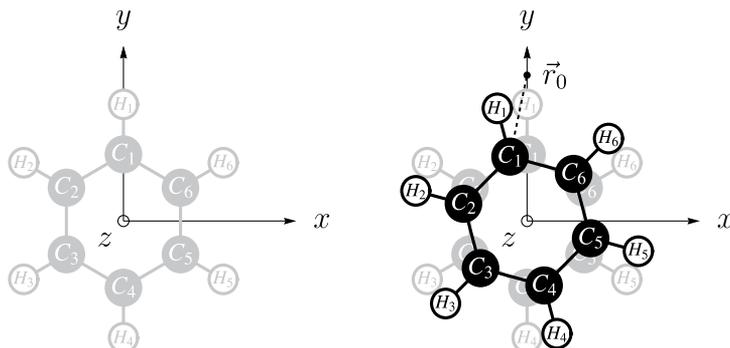


Fig. 3. Benzene molecule in the standard orientation (gray), with the carbon atom labeled  $C_1$  — the reference monitored point — on the  $y$ -axis. Resetting the dynamic monitored point is done by identifying the carbon atom closest to the reference monitored point as shown right where the present molecular orientation (black) is superimposed on the standard orientation (gray).

a real MC simulation of a molecular crystal. We also suggest a way to manage the resulting data, with the advantage of limited data storage in combination with unlimited simulation run lengths.

**Standard orientations. Reference monitored points.** As discussed above, for every type of molecule present in the simulated system, a standard orientation has to be defined first. Then, for every type of molecule, a family of points, related by symmetry operations of the molecule, has to be identified. For a molecule in the standard orientation, one of these points has to be defined as the reference monitored point. Let us recall the example of a crystal of benzene molecules: the standard orientation is shown on the left in Fig. 3, the family of symmetry-related points consists of the six C atoms, and the C atom with coordinates  $(x = 0, y > 0, z = 0)$  is the reference monitored point.

**Rotation matrices. Sequential runs.** It is convenient to put every molecule in its standard orientation at startup, and to describe the orientational state of molecule  $i$  by a rotation matrix  $R_i$ , so that at any time during the simulation, the current position  $\vec{r}'$  of an atom (or any point rotating along) of the molecule is obtained as  $\vec{r}' = R_i \vec{r}$ , where  $\vec{r}$  is the atom's position when the molecule is in the standard orientation. This does not prevent to initially put the molecules in random orientations, which is in fact recommended: instead of the identity matrix  $I$  every molecule then has an initial rotation matrix of the type  $R_{\vec{s}}(\alpha)$  [Eq. (16)] associated with it. During the simulation, if a rotational MC move with rotation matrix  $R_{\vec{s}}(\alpha)$  is accepted, the matrix  $R_i$  has to be updated to  $R_{\vec{s}}(\alpha)R_i$ . The matrix  $R_i$  is an array variable, and at the end of a program run of, say,  $Q$  steps, it is stored. When restarting the program, the matrices  $R_i$  are reloaded to regenerate the actual molecular orientations. We stress that it is not necessary to store the matrix  $R_i$  after every MC step, only at the end of a program run, to allow for a restart. This keeps data storage limited. We recommend to produce a long run of  $P$  MC steps by repeatedly running the MC program for  $Q$  MC steps. Here,  $Q$  is a divisor of  $P$  so that  $P = LQ$ , with  $P$ ,  $Q$  and  $L$  integers. The main advantages of restarting, valid in general when performing (MC) simulations, is that one can quickly study preliminary results, and that one can always resort to a previous, backed-up, simulation state in case of a machine crash. An additional advantage is that one can add features to the program without losing the simulation history. Finally, the memory required for storage can be easily reduced by reducing  $Q$ .

**Dynamic monitored points.** When equilibrium has set in, the sampling of OMSDs (and other quantities) can be started with. First, as explained above, the monitored points have to be reset. This is best done in a separate routine `reset`. For every molecule, the point closest to the reference monitored point has to be identified, and marked as the dynamic monitored point  $\tilde{r}_i$ . Typically, the result of this marking is an index referring to a specific atom (bond, ...) of molecule  $i$ . In Fig. 2, for the benzene example, the dynamic monitored point for molecule  $i$  has to be identified with  $C_6$  and for molecule  $j$  with  $C_4$ . OMSDs are calculated using the dynamic monitored point.

**OMSDs.** After each MC step  $p$ , the distance squared  $u_i(p)^2$  [Eq. (1)] is calculated and stored for every molecule  $i$ .<sup>1</sup> After the first run (let us call it the  $l = 1$  run) of  $Q$  steps after resetting the monitored points, the OMSDs are obtained as

$$\langle u_i^2 \rangle^{l=1} = \langle u_i^2 \rangle' = \frac{1}{Q} \sum_{p=1}^Q u_i(p)^2. \quad (17)$$

Here, the apostrophe refers to the last set of  $Q$  values while the superscript  $l$  refers to the total set of  $P = lQ$  simulation steps; the two sets obviously coincide for  $l = 1$ . This calculation is best done in a separate routine `process`. After a second run ( $l = 2$ ) of  $Q$  steps (without having reset the monitored points, otherwise it would again be a  $l = 1$  run), the OMSDs resulting from the second run only have to be calculated first:

$$\langle u_i^2 \rangle' = \frac{1}{Q} \sum_{p=1}^Q u_i(p)^2. \quad (18)$$

The overall OMSDs, combining the  $l = 1$  and  $l = 2$  runs, are then obtained as

$$\langle u_i^2 \rangle^{l=2} = \frac{1}{2} \langle u_i^2 \rangle' + \frac{1}{2} \langle u_i^2 \rangle^{l=1}. \quad (19)$$

In general, after  $L$  runs of  $Q$  MC steps each, and without resetting the monitored points between runs, the overall OMSDs, based on  $P = LQ$  MC simulation steps, read

$$\langle u_i^2 \rangle^{l=L} = \frac{1}{L} \langle u_i^2 \rangle' + \frac{L-1}{L} \langle u_i^2 \rangle^{l=L-1}, \quad (20)$$

where  $\langle u_i^2 \rangle'$  is the OMSD based on the last ( $L$ -th) run of  $Q$  steps only. In Fig. 4, the whole procedure is illustrated schematically. In this way, not the whole series of  $u_i(p)^2$  values has to be stored, but only the last sequence. Simulation runs of  $Q$  MC steps can be added infinitely, while the required amount for storage memory remains constant. The `reset` routine should

<sup>1</sup> In principle, a MC simulation can be performed without storing any sampled quantities. In the case of squared distances, for example, it suffices to add all values of  $u_i(p)^2$  during the simulation and to divide the resulting sum by the number of samples afterwards. When one does store sampled values like  $u_i(p)^2$ , one has the choice to store them in an array variable, or to write them to disk. Storing (the last set of  $Q$ ) samples has the advantage of having a "history" one can examine for debugging or physical understanding purposes. Particularly important is the evolution of the energy, which should not decrease when equilibrium has set in. Which programming style to use should depend on which optimization level one wants to achieve, which further depends on machine parameters (e.g. RAM and disk memory, I/O management, ...), and on personal taste. We leave it up to the reader to decide which style to use. To illustrate the concept of OMSDs in a clear way, we have opted for a description where sampled quantities are stored.

reset the counter  $l$  to 1. The `process` routine calculates  $\langle u_i^2 \rangle'$  and  $\langle u_i^2 \rangle^l$  and increments  $l$  to  $l + 1$ . Note that this way of updating average values based on the previous average value and on the last set of additionally calculated values can be applied to any sampled quantity (e.g. total energy, molecular translational mean-squared displacements, ...).

$$\begin{array}{c}
 \begin{array}{ccc}
 \underbrace{\boxed{1} \quad \dots \quad \boxed{Q}}_{l=1} & \underbrace{\boxed{1} \quad \dots \quad \boxed{Q}}_{l=2} & \dots \quad \underbrace{\boxed{1} \quad \dots \quad \boxed{Q}}_{l=L}
 \end{array} \\
 \langle u_i^2 \rangle' = \frac{1}{Q} \sum_{p=1}^Q u_i(p)^2 & \langle u_i^2 \rangle' = \frac{1}{Q} \sum_{p=1}^Q u_i(p)^2 & \langle u_i^2 \rangle' = \frac{1}{Q} \sum_{p=1}^Q u_i(p)^2 \\
 \underbrace{\langle u_i^2 \rangle^{l=1} = \langle u_i^2 \rangle'} \\
 \langle u_i^2 \rangle^{l=2} = \frac{1}{2} \langle u_i^2 \rangle' + \frac{1}{2} \langle u_i^2 \rangle^{l=1} \\
 \vdots \\
 \langle u_i^2 \rangle^{l=L} = \frac{1}{L} \langle u_i^2 \rangle' + \frac{L-1}{L} \langle u_i^2 \rangle^{l=L-1}
 \end{array}$$

Fig. 4. Scheme for calculating OMSDs based on  $P = LQ$  MC simulation steps arising from repeated runs of  $Q$  simulation steps.

#### 4. Case studies

After having introduced OMSDs and their practical implementation, we illustrate the usefulness of the concept with two examples involving rotating fullerenes: a 3D crystal containing  $C_{60}$  molecules, and a chain of one-dimensionally confined  $C_{70}$  molecules. Apart from applying the procedures discussed in the preceding sections, we add some extensions to the OMSD method along the way.

##### 4.1 Fullerene-cubane: orientational ordering in a molecular crystal of $C_{60}$ molecules

Fullerene-cubane, consisting of  $C_{60}$  and  $C_8H_8$  molecules, is a unique example of a molecular crystal combining highly symmetrical — icosahedral and cubic — molecules with stoichiometry 1:1 [Pekker et al. (2005)]. At room temperature the crystal lattice is face-centered cubic with the fullerene molecules rotating freely, while below  $T \approx 140$  K they adopt fixed orientations within an orthorhombic lattice. The cubane molecules do not rotate in both the high- and the low- $T$  phase, their faces are aligned with the crystal planes. Note that by “fixed orientations” the absence of molecular reorientations is understood; thermal librations are of course present (and are actually reproduced in MC simulations as will be seen from the OMSD values shown below).

In Verberck et al. (2009), an isothermal-isobaric ( $NpT$ -ensemble) MC simulation with simple Lennard-Jones pair interactions is reported, and the concept of OMSDs as outlined in the preceding sections is used to study the rotational behavior of the  $C_{60}$  (and  $C_8H_8$ ) molecules upon cooling from  $T = 300$  K to 50 K. Here we give an outline of how the OMSD treatment for the  $C_{60}$  molecules in fullerene-cubane is set up, and briefly discuss the results. For full details and an interpretation of the results in its physical context we refer to Verberck et al. (2009).

The standard orientation is shown in Fig. 5(a): three two-fold symmetry axes of the icosahedral  $C_{60}$  molecule coincide with the coordinate axes. The set of equivalent monitored points is chosen to be the set of the 30 double bonds (fusing hexagons of the soccer-ball shaped molecule). The double bond with coordinates  $(0, 0, z_d = 3.48 \text{ \AA})$  — for a molecule in the standard orientation, as the definition requires — serves as the reference monitored point [Fig. 5(b)]. Finally, the fixed point  $\vec{r}_0$  is chosen to coincide with the reference monitored point. We stress that the fixed and reference monitored points do not necessarily have to be the same.

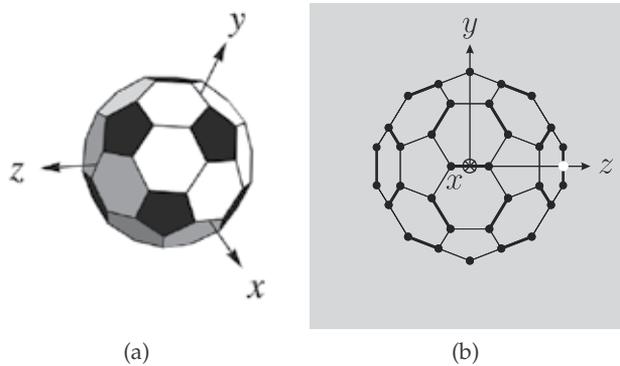


Fig. 5. (a) A  $C_{60}$  molecule, represented as a soccer ball with 12 pentagons (black) and 20 hexagons (white and shaded), in the standard orientation. (b) Projection onto the  $(y, z)$ -plane of a  $C_{60}$  molecule in the standard orientation. Double carbon-carbon bonds are shown thicker than single bonds. The double bond with coordinates  $(0, 0, z_d = 3.48 \text{ \AA})$  is marked by a white dot.

As stressed above, a reset to determine the dynamic monitored point  $\vec{r}_i$  is required before starting the sampling sequence of  $P$  MC steps. In the present example, the dynamic monitored point  $\vec{r}_i$  for molecule  $i$  is its double bond closest to the reference monitored point  $(0, 0, z_d = 3.48 \text{ \AA})$ .

Since the fixed point has vanishing  $x$ - and  $y$ -components, the analytical OMSDs for free 3D rotation and no rotation are  $r_i^2 + z_d^2$  and  $r_i^2 - 2r_i z_d \cos \theta_i + z_d^2$ , respectively (see Table 1), which reduce to

$$\langle u_i^2 \rangle_a = 2z_d^2 = 24.22 \text{ \AA}^2 \quad (\text{free 3D rotation}) \quad (21a)$$

and

$$\langle u_i^2 \rangle_a = 2z_d^2(1 - \cos \theta_i), \quad (\text{fixed orientation}) \quad (21b)$$

respectively, since  $r_i = |\vec{r}_i| = z_d$ . We recall that  $|\vec{r}_i|$  is indeed independent of the molecular index  $i$ , since the standard orientation and the reference monitored point are the same for every molecule.

As remarked already, working with only one fixed point  $\vec{r}_0$  is not enough to uniquely characterize the orientational regime. Here, for example, a fixed orientation with  $\theta_i = \pi/2$

results in the same OMSD value as for free 3D orientation. Note, however, that in this specific case, a fixed orientation with  $\theta_i = \pi/2$  is impossible since then another double bond would be closer to the reference monitored point, which would then instead have had to be set as the dynamic monitored point at reset. In fact, due to the high symmetry of the  $C_{60}$  molecule, implying high numbers of equivalent points (e.g. the 30 double bonds), the dynamic monitored point — in the case of no rotation or small librations — will always be close to the reference dynamic monitored point. Consequently, there is an upper limit for the highest value of  $\theta_i$ . Taking the atomic coordinates of the  $C_{60}$  molecule into account, the maximal value for  $\theta_i$  is about  $36^\circ$ . (For the benzene example of the previous section, which is easier to visualize, the maximal angle  $\theta_i$  is  $360^\circ/6 = 60^\circ$  — see Fig. 3.)

In Verberck et al. (2009) the average of the dynamic monitored point was used as an alternative fixed point  $\bar{r}_0$ . We will label the OMSD values resulting from such a type of “dynamic fixed point” as type II OMSDs, and call the OMSDs based on “static fixed points”, as in the original definition of Sect. 2, type I OMSDs, and use superscripts <sup>I</sup> and <sup>II</sup> for labeling the associated OMSD values. For type II OMSDs, the distances to be sampled are of the type

$$u_i(p)^2 = |\bar{r}_i(p) - \langle \bar{r}_i \rangle|^2 = [x_i(p) - \langle x_i \rangle]^2 + [y_i(p) - \langle y_i \rangle]^2 + [z_i(p) - \langle z_i \rangle]^2, \quad (22)$$

with

$$\langle \bar{r}_i \rangle = (\langle x_i \rangle, \langle y_i \rangle, \langle z_i \rangle) = \frac{1}{P} \sum_{p=1}^P \bar{r}_i(p). \quad (23)$$

At first sight, this requires to store all  $\bar{r}_i(p) = (x_i(p), y_i(p), z_i(p))$  values until the very end of the whole series of  $L = P/Q$  simulation runs of  $Q$  MC steps each, and that only then  $\langle \bar{r}_i \rangle$  and  $\langle u_i^2 \rangle = \frac{1}{P} \sum_{p=1}^P u_i(p)^2$  can be calculated. However, applying the usual “trick” for the evaluation of standard deviations in statistics,

$$\begin{aligned} \langle u_i^2 \rangle^{\text{II}} &= \frac{1}{P} \sum_{p=1}^P u_i(p)^2 \\ &= \frac{1}{P} \sum_{p=1}^P \left( [x_i(p) - \langle x_i \rangle]^2 + [y_i(p) - \langle y_i \rangle]^2 + [z_i(p) - \langle z_i \rangle]^2 \right) \\ &= \frac{1}{P} \sum_{p=1}^P (x_i(p)^2 - 2x_i(p)\langle x_i \rangle + \langle x_i \rangle^2 + y_i(p)^2 - 2y_i(p)\langle y_i \rangle + \langle y_i \rangle^2 \\ &\quad + z_i(p)^2 - 2z_i(p)\langle z_i \rangle + \langle z_i \rangle^2) \\ &= \langle x_i^2 \rangle - \langle x_i \rangle^2 + \langle y_i^2 \rangle - \langle y_i \rangle^2 + \langle z_i^2 \rangle - \langle z_i \rangle^2 \\ &= \langle r_i^2 \rangle - \langle x_i \rangle^2 - \langle y_i \rangle^2 - \langle z_i \rangle^2 \\ &= z_d^2 - \langle x_i \rangle^2 - \langle y_i \rangle^2 - \langle z_i \rangle^2, \end{aligned} \quad (24)$$

shows that storing the coordinates  $x_i(p)$ ,  $y_i(p)$  and  $z_i(p)$  during a simulation sequence of  $Q$  steps suffices. Indeed, the squares of the averages  $\langle x_i \rangle$ ,  $\langle y_i \rangle$  and  $\langle z_i \rangle$  are used to calculate the OMSD value  $\langle u_i^2 \rangle^{I=1} = \langle u_i^2 \rangle^I = \sum_{p=1}^Q u_i(p)^2$  based on the first simulation run of  $Q$  MC steps, which is then updated each time a sequence of  $Q$  MC steps is added according to the scheme

shown in Fig. 4. Again, the required computer memory remains constant (and can be reduced by reducing  $Q$ ), while segments of  $Q$  MC steps can be added ad libitum. Note that when using a static fixed point, i.e. a point  $\vec{r}_0 = (x_0, y_0, z_0)$  that does not “dynamically” depend on the simulation itself, the same approach involving the storing of the coordinates of  $\vec{r}_i$  can be used. Rather than calculating and storing the distance squared  $u_i(p)^2$  [Eq. (1)] after each MC step and averaging at the end, one can calculate the averages  $\langle x_i \rangle$ ,  $\langle y_i \rangle$  and  $\langle z_i \rangle$  and use the formula

$$\langle u_i^2 \rangle^I = z_d^2 - 2\langle x_i \rangle x_0 - 2\langle y_i \rangle y_0 - 2\langle z_i \rangle z_0 + r_0^2. \quad (25)$$

This approach can be considered when combining the two types of fixed points (static or dynamic) since for type II OMSDs  $\langle x_i \rangle$ ,  $\langle y_i \rangle$  and  $\langle z_i \rangle$  have to be calculated anyway. An important property of type II OMSDs is that they vanish in the case of fixed orientations, which follows trivially from  $\vec{r}_i(p) = \langle \vec{r}_i \rangle_a$ :

$$\langle u_i^2 \rangle_a^{II} = 0 \quad (\text{fixed orientation}). \quad (26)$$

Note that this property complements the orientation-dependent outcome of  $\langle u_i^2 \rangle_a^I$  [Eq. (21b)]. For free rotation,  $\langle \vec{r}_i \rangle_a = \vec{0}$ , so that

$$\langle u_i^2 \rangle_a^{II} = \langle |\vec{r}_i|^2 \rangle_a = z_d^2 = 12.11 \text{ \AA}^2 \quad (\text{free 3D orientation}). \quad (27)$$

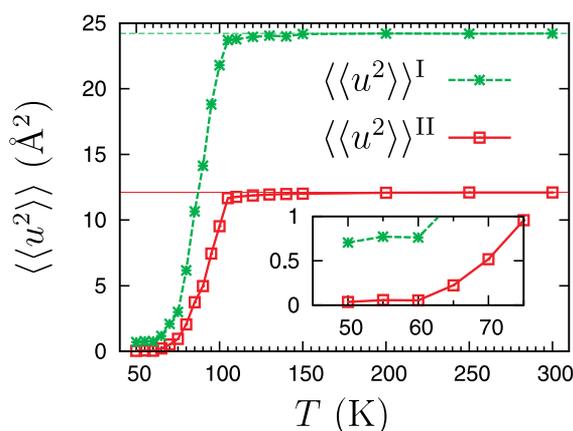


Fig. 6. Averaged type I and type II OMSD values  $\langle\langle u^2 \rangle\rangle^I$  and  $\langle\langle u^2 \rangle\rangle^{II}$  for  $C_{60}$  molecules in fullerene-cubane as a function of temperature  $T$ . The 300 K values correspond to the free rotation values of  $2z_d^2 = 24.22 \text{ \AA}^2$  and  $z_d^2 = 12.11 \text{ \AA}^2$ , respectively (shown by horizontal lines).

The type I and type II OMSD values resulting from the simulation, averaged over all molecules, written as  $\langle\langle u^2 \rangle\rangle^I$  and  $\langle\langle u^2 \rangle\rangle^{II}$  [see Eq. (3)], respectively, are shown in Fig. 6 as a function of temperature  $T$ . They nicely show a transition from freely rotating to orientationally frozen molecules. Indeed, the 300 K values match the exact analytical values of  $z_d^2$  and  $2z_d^2$ , while at 50 K,  $\langle\langle u^2 \rangle\rangle^{II} \approx 0$ , implying fixed molecular orientations. The small deviations of  $\langle\langle u^2 \rangle\rangle^{II}$  from zero (see Inset of Fig. 6) are a consequence of thermally induced librations. The transition covers the temperature range  $65 \text{ K} \lesssim T \lesssim 110 \text{ K}$ , which can be interpreted as a

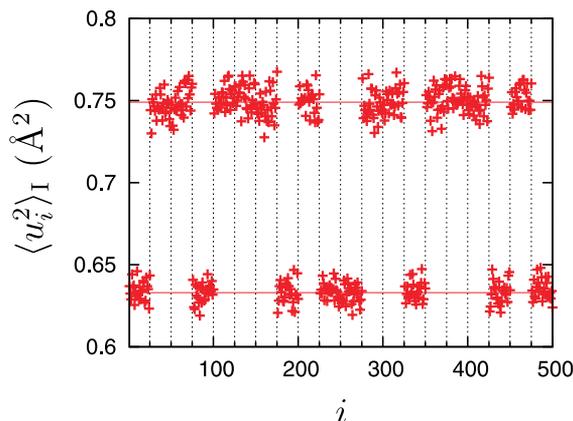


Fig. 7. Molecular  $\langle u_i^2 \rangle_I$  OMSD values of the 500  $C_{60}$  molecules in the simulation box, at  $T = 50$  K. Two groups of molecules, with values around  $0.633 \text{ \AA}^2$  and  $0.749 \text{ \AA}^2$  (shown by horizontal lines), are clearly distinguishable.

slow freezing of the rotational motion, and could be observed experimentally as a continuous diminishing of diffuse x-ray scattering.

The finite low- $T$  values of  $\langle \langle u^2 \rangle \rangle_I$  correspond to specific fixed molecular orientations. It is, however, wrong to a priori assume that all molecules adopt the same orientation. The individual molecular  $\langle u_i^2 \rangle_I$  values have to be examined to conclude upon the precise molecular orientations. In Fig. 7, the  $T = 50$  K  $\langle u_i^2 \rangle_I$  values for the 500  $C_{60}$  molecules in the simulation box are shown. Interestingly, the values can be divided into two groups, with average values  $\langle u_i^2 \rangle_I \approx 0.633 \text{ \AA}^2$  and  $\langle u_i^2 \rangle_I \approx 0.749 \text{ \AA}^2$ , corresponding to two classes of molecular orientations. The details of the determination of the molecular orientations resulting in Fig. 7 fall beyond the scope of the present Chapter; for a complete analysis we refer to Verberck et al. (2009). We do point out, however, that it turns out useful to keep track of the averages  $\langle x_i \rangle$ ,  $\langle y_i \rangle$  and  $\langle z_i \rangle$  since they help to deduce the molecules' orientations. Indeed, knowledge of  $\langle \vec{r}_i \rangle$  already fixes the  $i$ -th molecule's average orientation up to a rotation about the vector  $\langle \vec{r}_i \rangle$ . Note that this requires almost no extra programming and memory since averages of the dynamic monitored points' coordinates have already been included in the implementation of type II OMSDs [cfr. Eq. (24)].

#### 4.2 Nanopeapods: orientational behaviour in one-dimensional chains of $C_{70}$ molecules

Our second example illustrating the concept of OMSDs is a chain of  $C_{70}$  molecules encapsulated in a CNT. Such a system falls into the category of so-called nanopeapods, nanotubes filled with atom or molecules. Historically, the insertion of  $C_{60}$  molecules was reported first [Smith et al. (1998)], but many other peapod systems have been synthesized and investigated by now. For a review, we refer to Monthieux (2002).

One of the interesting properties of  $C_{70}@CNT$  systems is the dependence of the molecule's orientation on the tube radius  $R$ : for a small radius, the molecules adopt lying orientations [Fig. 8(a)] while for larger  $R$ , molecules adopt standing orientations [Fig. 8(b)]. Obviously, this is a consequence of the molecule's geometry and its van der Waals interaction with the surrounding nanotube. Recently, a canonical-ensemble (NVT) MC simulation of  $C_{70}$  peapods in a CNT modelled as a homogeneous carbonic cylinder was carried out in order to study the temperature and the radius dependence of the molecular motion of the one-dimensionally

confined fullerene molecules [Verberck et al. (2011)]. In particular, OMSDs were used to characterize the molecules' rotational behavior. Here, as in the previous subsection, we discuss how the OMSD analysis is set up, and point to possible additions.

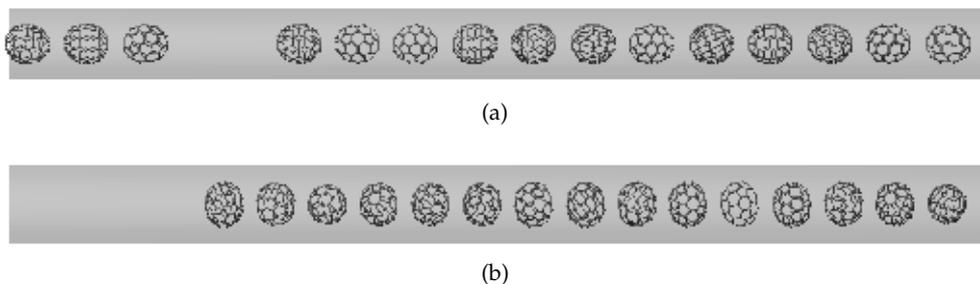


Fig. 8. Carbon nanotubes filled with C<sub>70</sub> molecules. (a) For a tube radius of 6.5 Å, the molecules adopt lying orientations (the molecule's long axis parallel to molecule's long axis). (b) For a radius of 7.3 Å, the molecules adopt standing orientations (at sufficiently low temperatures).

As a first step, a standard orientation for the investigated molecules has to be defined. The C<sub>70</sub> standard orientation is shown in Fig. 9; it is a lying orientation. Next, a set of equivalent monitored points has to be identified. A convenient choice is the pair of top and bottom pentagons' centers  $(0, 0, \pm z_p)$ , with  $z_p = 3.99$  Å. The point with the positive z-coordinate is chosen as the reference monitored point (Fig. 9).

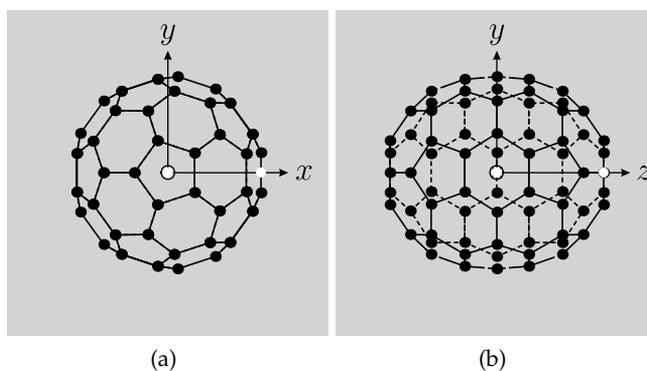


Fig. 9. Projections of a C<sub>70</sub> molecule in the standard orientation on the (a)  $(x, y)$ - and the (b)  $(y, z)$ -plane. The two reference monitored points  $(0, 0, z_t = 3.99$  Å) and  $(x_b = 3.49$  Å,  $0, 0)$  are marked by white dots.

Rather than working with type I and type II OMSDs for one fixed point, as in the example of the previous subsection, only type II OMSDs were used in Verberck et al. (2011), but for two monitored points. We therefore introduce superscripts  $II,1$  and  $II,2$ . The second set of monitored points is the ring of five bonds in the equatorial plane ( $z = 0$ ) of the C<sub>70</sub> molecule in the standard orientation; the point  $(x_b, 0, 0)$  with  $x_b = 3.49$  Å is set as the second reference monitored point (Fig. 9). For free 3D rotations, we have — see the previous subsection —

$\langle u_i^2 \rangle_a^{\text{II}} = \langle |\vec{r}_i|^2 \rangle_a$  so that the type II OMSDs read

$$\langle u_i^2 \rangle_a^{\text{II},1} = z_t^2 = 15.90 \text{ \AA}^2 \quad (\text{free 3D orientation}), \quad (28a)$$

$$\langle u_i^2 \rangle_a^{\text{II},2} = x_b^2 = 12.15 \text{ \AA}^2 \quad (\text{free 3D orientation}). \quad (28b)$$

For fixed orientations, the OMSDs vanish:

$$\langle u_i^2 \rangle_a^{\text{II},1} = 0 \quad (\text{no rotation}), \quad (29a)$$

$$\langle u_i^2 \rangle_a^{\text{II},2} = 0 \quad (\text{no rotation}). \quad (29b)$$

Another case of interest is free molecular rotation about the z-axis (cfr. Sect. 2). Using the appropriate average for this case, Eq. (8), one easily obtains

$$\langle \vec{r}_i \rangle_a = \langle (r_i \sin \theta_i \cos \phi_i, r_i \sin \theta_i \sin \phi_i, r_i \cos \theta_i) \rangle_a = (0, 0, r_i \cos \theta_i), \quad (30)$$

so that, using Eq. (24),

$$\langle u_i^2 \rangle_a^{\text{II}} = r_i^2 (1 - \cos^2 \theta_i), \quad (31)$$

resulting in

$$\langle u_i^2 \rangle_a^{\text{II},1} = z_t^2 (1 - \cos^2 \theta_i) \quad (\text{free z-orientation}), \quad (32a)$$

$$\langle u_i^2 \rangle_a^{\text{II},2} = x_b^2 (1 - \cos^2 \theta_i) \quad (\text{free z-orientation}). \quad (32b)$$

The polar angle of the first dynamic monitored point equals 0 for lying and  $\pi/2$  for standing molecules, respectively, so that

$$\langle u_i^2 \rangle_a^{\text{II},1} = 0 \quad (\text{lying, free z-orientation}), \quad (33a)$$

$$\langle u_i^2 \rangle_a^{\text{II},1} = z_t^2 \quad (\text{standing, free z-orientation}). \quad (33b)$$

For the second dynamic monitored point  $\theta_i = \pi/2$  for any lying orientation. For standing orientations, the value of  $\theta_i$  can adopt any value between 0 and  $\pi$ , depending on the molecule's orientation with respect to its own long axis. Hence

$$\langle u_i^2 \rangle_a^{\text{II},2} = x_b^2 \quad (\text{lying, free z-orientation}), \quad (34a)$$

$$\langle u_i^2 \rangle_a^{\text{II},2} = x_b^2 (1 - \cos^2 \theta_i) \quad (\text{standing, free z-orientation}). \quad (34b)$$

It can be shown that in the case of a standing molecule, rotating freely about the z-axis and freely about its long axis (i.e. spinning freely),  $\langle \cos^2 \theta_i \rangle = 0$  for the second dynamic monitored point, so that

$$\langle u_i^2 \rangle_a^{\text{II},2} = x_b^2 \quad (\text{standing, free z-orientation, free spinning}). \quad (35)$$

The various cases are summarized in Table 2.

It follows that only lying molecules rotating freely about the z-axis and absence of molecular rotation can be unambiguously inferred from the pair  $(\langle u_i^2 \rangle_a^{\text{II},1}, \langle u_i^2 \rangle_a^{\text{II},2})$ . Free 3D rotations and free rotations about the z-axis in combination with free spinning of the molecule about its long axis result in the same OMSDs. The case of free z-rotation of a standing molecule, not spinning about its long axis, and with a fixed polar angular value of  $\theta_i = \pi/2$  for the second

			$\langle u_i^2 \rangle_a^{\text{II},1}$	$\langle u_i^2 \rangle_a^{\text{II},2}$
free 3D rotation			$z_t^2 = 15.90 \text{ \AA}^2$	$x_b^2 = 12.15 \text{ \AA}^2$
free z-rotation	lying		0	$x_b^2$
	standing	free spinning	$z_t^2$	$x_b^2$
		no spinning	$z_t^2$	$x_b^2(1 - \cos^2 \theta_i)$
no rotation			0	0

Table 2. Analytical OMSD values  $\langle u_i^2 \rangle_a^{\text{II},1}$  and  $\langle u_i^2 \rangle_a^{\text{II},2}$  for reference monitored points  $(0, 0, z_t)$  and  $(x_b, 0, 0)$ , respectively, for several cases of interest.

dynamic monitored point, also results in the same pair of OMSDs. This accidental coincidence can be easily resolved, though, by considering a third monitored point (e.g. any other bond in the equatorial belt of the  $C_{70}$  molecule, see Fig. 9). On the other hand, the coincidence for free 3D rotation and free z-rotation plus free spinning is hard to lift using OMSDs only. A simpler solution, described in Verberck et al. (2011) is to monitor the z-coordinate of the first dynamic monitored point. Indeed, for free 3D rotation, its value should be uniformly distributed in the interval  $[-z_t, z_t]$ , while for a permanently standing molecule (regardless of whether it rotates about the z-axis and/or spins about its long axis) its value should be 0. To extract the distribution of z-coordinate values in a MC simulation, it is necessary to make a histogram. This can be either done on-the-fly, by incrementing the bin count of the bin corresponding to the current z-coordinate, or in the `process` routine if the  $z_i$  values of the dynamic monitored point are stored during the simulation.

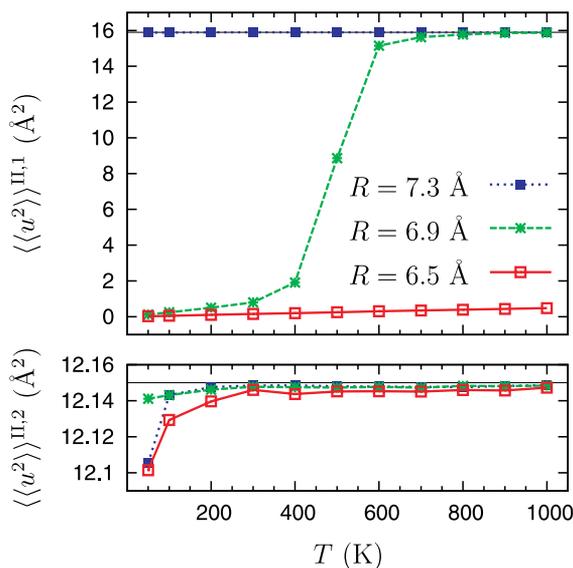


Fig. 10. Averaged type II OMSD values  $\langle\langle u^2 \rangle\rangle^{\text{II},1}$  and  $\langle\langle u^2 \rangle\rangle^{\text{II},2}$  for  $C_{70}$  nanopeapods with three different radii  $R$ , as a function of temperature  $T$ . The free rotation values of  $z_t^2 = 15.90 \text{ \AA}^2$  and  $x_b^2 = 12.15 \text{ \AA}^2$  are shown by horizontal lines.

In Fig. 10, the OMSDs  $\langle u_i^2 \rangle_a^{\text{II},1}$  and  $\langle u_i^2 \rangle_a^{\text{II},2}$ , averaged over all (15) molecules, resulting from the MC simulation are shown for three different radii, as a function of temperature. The smallest tube radius,  $R = 6.5 \text{ \AA}$ , features lying molecules, freely rotating about the long axis of the

tube ( $z$ -axis). At low temperatures, judging from the values of  $\langle u_i^2 \rangle_a^{II,2}$  being slightly smaller than  $x_b^2$ , the free character of the rotations is only almost reached, which can be attributed to intermolecular interactions favoring specific mutual orientations. At high temperatures, truly free rotation has been achieved, but now the small deviation of  $\langle u_i^2 \rangle_a^{II,1}$  from zero suggests thermal fluctuations — very small tilts away from the ideal lying orientation.

For the highest radius,  $R = 7.3 \text{ \AA}$ , the OMSD values imply that molecules rotate freely, either three-dimensionally or around the  $z$ -axis (with spinning). The procedure involving histograms of the  $z$ -coordinates of the first dynamic monitored point mentioned above reveals a transition from standing molecules at low temperatures to complex 3D rotations at higher temperatures. For details, we refer to Verberck et al. (2011).

The intermediate radius  $R = 6.9 \text{ \AA}$  shows a transition from quasi-freely (cfr.  $R = 6.5 \text{ \AA}$ )  $z$ -rotating lying molecules to — as follows from the  $z$ -coordinate histogram analysis — a complex pattern of 3D molecular rotations.

## 5. Summary

In the present Chapter, we have introduced a method to quantify molecular rotational motion in molecular crystals, and have shown how to embed it in a typical MC simulation. The key concept is that of OMSDs, a rotational analogue of translational mean-squared displacements. By carefully choosing a point that rotates along with a molecule, following it during the simulation, and calculate its distance squared to a fixed point, a measure for the “degree of rotation” of the molecule is obtained. We have shown how to properly set up this procedure to avoid biased results due to the initial equilibrating phase and due to molecular symmetry. Also, we discussed practical issues required for the efficient implementation of the OMSD technique.

Two examples of MC simulations where OMSDs were used — a 3D molecular crystal consisting of  $C_{60}$  molecules and a 1D arrangement of  $C_{70}$  molecules in a nanotube — were covered. The examples show the power of the OMSD method. By combining two types of OMSDs and using more than one monitored point, it is possible to deduce the rotational regimes of the molecules in a simulation box. Typical regimes are free 3D rotations, free rotations about an axis, or no rotations. In the latter case, the examination of OMSD values helps to determine the actual molecular orientations. The examples also provide hints to resolve occasional ambiguities. Indeed, some types of rotational motion can be indistinguishable within a certain set of OMSDs. In these cases, looking at average values or histograms of coordinates of the monitored point(s) — which requires minimal additional programming since these are parameters already required for the calculation of OMSDs — or extending the set of OMSDs can resolve the problem.

For completeness, we point out that the procedures described in the preceding sections can be easily extended to the case of a crystal of several types of molecules. It simply suffices to define a standard orientation, monitored points (reference and dynamic) and fixed points for each type of molecule, and to process the several molecular sublattices separately.

In summary, the OMSD method provides a way to map the huge number of variables (atomic coordinates) in a MC simulation of a molecular crystal onto a very small set of quantities completely describing the molecule’s rotational motion.

## Acknowledgements

The author has benefited from collaborations with G.A. Vliegthart, G. Gompper, J. Cambedouzou and P. Launois. In addition, the author acknowledges useful discussions with K.H. Michel and M. Ripoll.

## 6. References

- Allen, M.P. & Tildesley, D.J. (1987). *Computer Simulation of Liquids*, Oxford University Press, ISBN 0-19-855645-4, Oxford
- Frenkel, D. & Smit, B. (2002). *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd Ed., Academic Press, ISBN 0-12-267351-4, San Diego
- Lynden-Bell, R.M. & Michel, K.H. (1994). Translation–rotation coupling, phase transitions, and elastic phenomena in orientationally disordered crystals. *Reviews of Modern Physics*, 66, 3 (July 1994) (721-762), ISSN 0034-6861
- Miles, R.E. (1965). On random rotations in  $\mathbb{R}^3$ . *Biometrika*, 52, 3/4 (December 1965) (636-639), ISSN 0006-3444
- Monthioux, M. (2002). Filling single-wall carbon nanotubes. *Carbon*, 40, 10 (August 2002) (1809-1823), ISSN 0008-6223
- Pekker, S.; Kováts, É.; Oszlányi, G.; Bényei, G.; Klupp, G.; Bortel, G.; Jalsovszky, I.; Jakab, E.; Borondics, F.; Kamarás, K.; Bokor, M.; Kriza, G.; Tompa, K. & Faigel, G. (2005). Rotor–stator molecular crystals of fullerenes with cubane. *Nature Materials*, 4, 10 (September 2005) (764-767), ISSN 1476-1122
- Smith, B.W.; Monthioux, M. & Luzzi, D.E. (1998). Encapsulated  $C_{60}$  in carbon nanotubes. *Nature*, 396, 6709, (26 November 1998) (323-324), ISSN 0028-0836
- Verberck, B.; Vliegthart, G.A. & Gompper, G. (2009). Orientational ordering in solid  $C_{60}$  fullerene-cubane. *The Journal of Chemical Physics*, 130, 15 (21 April 2009) (154510.1-14), ISSN 0021-9606
- Verberck, B.; Cambedouzou, J.; Vliegthart, G.A.; Gompper, G. & Launois, P. (2011). Molecular motion in  $C_{70}$ @SWCNT nanopeapods: a Monte Carlo study. Accepted for publication in *Carbon*, ISSN 0008-6223
- Vesely, F.Z. (1982). Angular Monte Carlo Integration Using Quaternion Parameters: A Spherical Reference Potential for  $CCl_4$ . *Journal of Computational Physics*, 47, 2 (August 1982) (291-296), ISSN 0021-9991

# Finite-time Scaling and its Applications to Continuous Phase Transitions

Fan Zhong

*State Key Laboratory of Optoelectronic Materials and Technologies  
School of Physics and Engineering, Sun Yat-sen University, Guangzhou 510275  
People's Republic of China*

## 1. Introduction

Monte Carlo methods of numerical simulations play an important role in studying phase transitions in general and critical phenomena in particular in statistical physics (Amit & Martin-Mayor, 2005; Binder & Heermann, 1988; Landau & Binder, 2005; Newman & Barkema, 1999). A hallmark of the critical phenomena that a system exhibit in the vicinity of a second-order phase transition, or in the modern classification, continuous phase transition (Fisher, 1967) is its diverging correlation length (Amit & Martin-Mayor, 2005; Cardy, 1996; Ma, 1976; Stanley, 1971). This length scale renders at first sight numerical simulations useless because they are inevitably carried out on systems of finite sizes that are thus smaller than the correlation length and thus cannot probe the bulk behavior of the system under considered. Moreover, real phase transitions occur only in the thermodynamic limit.

Yet, the idea of finite-size scaling has turned this nuisance into a blessing and the method based on it has become a routine to extract critical properties from numerical simulations of finite systems (Amit & Martin-Mayor, 2005; Cardy, 1988; Fisher & Ferdinand, 1967; Fisher & Barber, 1972; Gasparini et al., 2008; Landau & Binder, 2005; Privman, 1990). Under the assumption that upon a renormalization-group transformation of a length rescaling of factor  $b$ , the coupling constants of a finite system transform in the same way as in the thermodynamics limit (Brézin, 1982; Brézin & Zinn-Justin, 1985), the singular part of the free energy of the system then transforms as

$$F(\tau, H, L^{-1}) = b^{-d} F(\tau b^{1/\nu}, H b^{\beta\delta/\nu}, L^{-1} b), \quad (1)$$

where  $\delta$ ,  $\beta$ , and  $\nu$  are critical exponents,  $L$  is a characteristic length scale of the system,  $d$  the spatial dimensionality,  $H$  the external magnetic field (we shall use the terminology of magnetism throughout), and the reduced temperature  $\tau = T - T_c$  with  $T_c$  being the critical temperature. As a result, one arrives at the finite-size scaling ansatz for the free energy

$$F(\tau, H, L^{-1}) = L^{-d} f(\tau L^{1/\nu}, H L^{\beta\delta/\nu}), \quad (2)$$

where  $f$  is a scaling function. We have neglected possible dimensional factors for conciseness hereafter. Appropriate differentiations of Equation (2) then give rise to corresponding scaling

forms for the magnetization  $M$ , the susceptibility  $\chi$ , and the specific heat  $C$  as

$$M(\tau, L) = L^{-\beta/\nu} f_1(\tau L^{1/\nu}), \quad (3a)$$

$$\chi(\tau, L) = L^{\gamma/\nu} f_2(\tau L^{1/\nu}), \quad (3b)$$

$$C(\tau, L) = L^{\alpha/\nu} f_3(\tau L^{1/\nu}) \quad (3c)$$

using the scaling laws or relations

$$\alpha = 2 - d\nu, \quad (4a)$$

$$\alpha + 2\beta + \gamma = 2, \quad (4b)$$

$$\beta\delta = \beta + \gamma, \quad (4c)$$

where  $\alpha$  and  $\gamma$  are critical exponents and the  $f_s$  including those that will appear later are all scaling functions. In terms of the infinite system correlation length  $\xi_\infty$  that diverges at  $T_c$  as

$$\xi_\infty \propto |\tau|^{-\nu}, \quad (5)$$

the argument of  $f_s$  in Equations (3) is proportional to  $L/\xi_\infty$  that governs the finite-size behavior; for small  $L/\xi_\infty$ , finite-size scaling appears in which  $L$  is a relevant length scale, while large  $L/\xi_\infty$  is the thermodynamic limit in which equilibrium behavior shows and  $L$  is irrelevant. Note that all the critical exponents assume their infinite-lattices values due to the aforementioned assumption (Brézin, 1982; Brézin & Zinn-Justin, 1985). Consequently, measuring the observables for a series of  $L$  can then determine the corresponding exponent ratios and finally the critical exponents themselves, the pitch of the critical properties, from the pure power laws emerged exactly at  $T_c$  or  $\tau = 0$  at which  $f_s$  are assumed to be analytic. In fact, for too small systems sizes and temperatures too far away from  $T_c$ , corrections to scaling (Wegner, 1972) have to be taken into account. Nevertheless, delicate methods have been developed for extracting critical exponents as well as  $T_c$  (Amit & Martin-Mayor, 2005; Landau & Binder, 2005).

A sequence of Monte Carlo updates may be interpreted as a discrete Markov process (Glauber, 1963; Landau & Binder, 2005; Müller-Krumbhaar & Binder, 1973). Consequently, Monte Carlo simulations can also be applied to study time-dependent dynamic behavior, though usually studied is stochastic relaxational dynamics instead of 'true dynamics' in which the dynamics is determined by the equations of motion derived from a Hamiltonian. Yet, the stochastic dynamics for the kinetic Ising model with local spin dynamics as realized in the single-site Metropolis algorithm (Metropolis et al., 1953), for instance, is believed to fall into the same universality class as that governed by the time-dependent Ginzburg-Landau equation (Hohenberg & Halperin, 1977). Dynamic critical phenomena (Cardy, 1996; Ferrell et al., 1967; Folk & Moser, 2006; Halperin & Hohenberg, 1967; Hohenberg & Halperin, 1977; Ma, 1976) are also accompanying with a divergent correlation time  $t_{eq}$  which diverges with the correlation length  $\xi_\infty$  as

$$t_{eq} \propto \xi_\infty^z \quad (6)$$

with a new dynamical critical exponent  $z$  dynamic finite-size scaling (Suzuki, 1977) can be obtained by formally incorporating the time argument  $t$  in Equation (1), giving rise to

$$M(\tau, H, t, L^{-1}) = b^{-\beta/\nu} M(\tau b^{1/\nu}, H b^{\beta\delta/\nu}, t b^{-z}, L^{-1} b) \quad (7)$$

after a derivative with  $H$ . As a result, the finite-size scaling form of the order parameter, Equation (3a), say, now becomes

$$M(\tau, t, L) = L^{-\beta/\nu} f_{1t}(\tau L^{1/\nu}, tL^{-z}), \quad (8)$$

which implies a dynamic finite-size scaling form for the correlation time

$$t_L = L^z f_{2t}(\tau L^{1/\nu}). \quad (9)$$

Therefore, at the criticality,

$$t_L \propto L^z \quad (10)$$

in the asymptotic region of large time, large size, and small  $\tau$ . This is again a standard method to estimate  $z$ , though when the asymptotic region is reached is not easy to determine (Landau & Binder, 2005; Wansleben & Landau, 1991).

However, actual simulations can only be performed inevitably in a limited time for large system sizes. So, one encounters a situation that is similar to the static case: In order to have good estimates of  $z$ , one needs to wait for a long time that is longer than  $t_{eq}$  to enter the asymptotic region similar to the static case in which one needs a large system size that is bigger than  $\zeta$ . In fact, even in the static case, one also needs to wait a similar long time to the dynamic case in order for the system to equilibrate. This is in fact the issue of critical slowing down. Nevertheless, finite-size scaling has efficiently helped the static case to overcome the limited-size problem. It is then quite surprised that a finite-time scaling was elusive for nearly forty years except for several not-well-noticed work on disordered systems (Hukushima & Nemeto, 1995; Shima & Nakayama, 1998; 1999; Shinomoto & Kabashima, 1991). Recently, we realized that the linearly driving method we had been applying to study both first-order (Zhang & Zhong, 1996; Zhong, 2002; Zhong & Chen, 2005; Zhong et al., 1994; 1995; 1998; Zhong & Zhang, 1995a,b; 1997) and continuous phase transitions (Fan & Zhong, 2007; 2009; Zhong, 2006; Zhong & Xu, 2005) just offered a realization of such a scaling (Gong et al., 2010; Huang et al., 2010). The method provides an external effective time scale that is inversely proportional to the rate  $R$  of the driving by which either an external field or the temperature that varies linearly with time was applied to a system near its criticality. Because of this finite time scale, the system evolves according to the driving rather than by itself, which takes a long time near its criticality. As a consequence, both static and dynamic scaling behavior can be probed effectively without suffering from critical slowing down. In addition, this time scale is readily manipulable not only in simulations but also in experiments (Zhang & Zhong, 1996) and thus serves as the temporal analogue of the finite size scale, though the latter may not be so obtainable experimentally (Gasparini et al., 2008). We shall review the theory and applications of the finite-time scaling in this chapter. However, before entering into the details, we would like to make some remarks.

First, for the usual Monte Carlo simulations of phase transitions such as those on the Ising model with the usual Metropolis algorithm (Metropolis et al., 1953), a direct analogue of the finite-size scaling by measuring observables at a series of time may not work because the system needs sufficiently long time to sample its configuration space in order for the observables to be measured correctly.

Second, there exist scalings with the time or its Fourier transform frequency. However, these are not the kinds of finite-time scaling in the spirit of the finite-size scaling. The central distinction is that in finite-time scaling there is a driving that imposes on the system a finite

time scale that restricts its natural evolution and thus its scaling and that is controllable in close analogue to the external length scale in finite-size scaling.

For example, in the short-time critical dynamics (Zheng, 1998), there is a new independent initial slip exponent  $\theta$  associated with a small initial magnetization  $m_0$  (Janssen et al., 1989). The dynamic transformation law for  $M$  in the absence of  $H$  is (Li et al., 1995; Zheng, 1998)

$$M(\tau, t, m_0, L) = b^{-\beta/\nu} M(\tau b^{1/\nu}, t b^{-z}, b^{x_0} m_0, L^{-1} b), \quad (11)$$

or

$$M(\tau, t, m_0, L) = t^{-\beta/\nu z} f_{3t}(\tau t^{1/\nu z}, m_0 t^{x_0/z}, L^{-1} t^{1/z}) \quad (12)$$

by setting  $b = t^{1/z}$ , where  $x_0 = \theta + \beta/\nu$ , because for a sufficiently large lattice and small  $m_0$  at  $\tau = 0$ ,  $M \propto t^\theta$  in the initial stage from Equation (12). In addition, one may make a temporal Fourier transformation to the dynamics. A transformation law in terms of the frequency  $\omega$  instead of  $t$  similar to Equation (7) can be written, leading then to

$$M(\tau, \omega, L) = \omega^{\beta/\nu z} f_{4t}(\tau \omega^{-1/\nu z}, L^{-1} \omega^{-1/z}). \quad (13)$$

One may regard Equations (12) and (13) as examples of finite-time scaling. However,  $t$  and its Fourier transform  $\omega$  are natural evolution time of the system and cannot be varied in contrast to  $L$  in Equation (8) and  $R$  in Equations (34) and (47) below.

In the following, we shall first review briefly the theory of finite-time scaling (Section 2), then summarize the current methods to extract critical properties using finite-time scaling (Section 3), followed by a summary of the results obtained with them in continuous transitions of pure and disordered systems (Section 4). Finally, discussions on the merits and shortcomings of the methods and future studies as well as conclusions are presented in Section 5.

## 2. Theory of finite-time scaling

In this section, we shall first briefly review in Section 2.1 the renormalization-group theory of finite-time scaling both for a field driving (Section 2.1.1) and for a temperature driving (Section 2.1.2) to justify the scaling (Gong et al., 2010; Zhong, 2006). Then, we shall study crossovers in Section 2.2 and corrections to scaling in Section 2.3 and discuss the combined effects of both finite time and finite size by taking into account the latter in Section 2.4.

### 2.1 Renormalization-group theory of finite-time scaling

Consider the model with the following Ginzburg-Landau functional of a  $\varphi^4$  theory in an external field  $H$ ,

$$F[\varphi] = \int d\mathbf{r} \left\{ \frac{1}{2} \tau \varphi^2 + \frac{1}{4!} g \varphi^4 + \frac{1}{2} (\nabla \varphi)^2 - H \varphi \right\}, \quad (14)$$

where  $g$  is a coupling constant and  $\tau$  is proportional to the temperature distance from the mean-field  $T_c$ . Its dynamics is governed by the Langevin equation for the scalar non-conserved order parameter  $\varphi$ ,

$$\frac{\partial \varphi}{\partial t} = -\lambda \frac{\delta F[\varphi]}{\delta \varphi} + \xi \quad (15)$$

with a Gaussian white noise  $\zeta$  satisfying

$$\langle \zeta(\mathbf{r}, t) \rangle = 0, \quad \langle \zeta(\mathbf{r}, t) \zeta(\mathbf{r}', t') \rangle = 2\lambda \delta(\mathbf{r} - \mathbf{r}') \delta(t - t'), \quad (16)$$

where  $\lambda$  is a kinetic coefficient. This is the Model A (Hohenberg & Halperin, 1977), which falls in the same universality class as the kinetic Ising model with local spin dynamics. We shall consider two driven non-equilibrium situations in which one starts with a sufficiently ordered state and increases linearly either the external field  $H = Rt$  or the temperature  $\tau = Rt$  with a small rate constant  $R$  across the critical point. As the critical point lies at  $\tau = 0$  and  $H = 0$ , our choice of initial conditions means  $t$  is in fact  $t - t_c$  with  $t_c$  the time at the critical point. For simplicity, however, we shall still use  $t$  below as the shift of initial point makes no difference in the linear driving, which makes in fact this driving superior to others including the usual sinusoidal driving (Gong et al., 2010).

In order to use systematic field-theoretic methods, we recast the dynamics into an equivalent field theory with a dynamic functional (Janssen, 1992),

$$I[\varphi, \tilde{\varphi}] = \int d\mathbf{r} dt \left\{ \tilde{\varphi} \left[ \dot{\varphi} + \lambda(\tau - \nabla^2)\varphi + \frac{1}{3!} \lambda g \varphi^3 - \lambda H \right] - \lambda \tilde{\varphi}^2 \right\} \quad (17)$$

by introducing an auxiliary response field  $\tilde{\varphi}$  (Martin et al., 1973). Expectation values can then be obtained by taking appropriate derivatives of the generating functional

$$W[h, \tilde{h}] = \ln \int D(\varphi, \tilde{\varphi}) \exp[-I[\varphi, \tilde{\varphi}] + \int d\mathbf{r} dt (h\varphi + \tilde{h}\tilde{\varphi})] \quad (18)$$

with respect to the external sources  $h$  and  $\tilde{h}$  that conjugate respectively to  $\varphi$  and  $\tilde{\varphi}$ .

Accordingly to the field theoretical formulation of the renormalization-group theory, the critical exponents are associated with the renormalization factors  $Z$ s that cure the divergences in the theory (Amit & Martin-Mayor, 2005; Zinn-Justin, 1996). One notices that since the variation of  $T$  and  $H$  is spatially uniform and depends linearly on time with a small  $R$ , no new divergence except the extrinsic one at  $t \rightarrow \infty$  or  $\omega \rightarrow 0$  in frequency domain is generated. As a result, no new  $Z$  besides the usual  $\varphi^4$ -theory ones has to be introduced, except the possible initial slip (Janssen et al., 1989) which does not contribute however because the transition is independent of the initial condition when we start with a driving sufficiently far away from the critical point. To deal with the time-dependent external probes, we perform the renormalization at the critical point at which  $\tau$  and  $H$  vanish, and then make an expansion about the critical theory (Amit & Martin-Mayor, 2005; Weinberg, 1973; Zinn-Justin, 1996) by taking as insertions the deviations arising from the driving away from that point. In this way, the time dependent external probes can be naturally accounted for. Moreover, the renormalization at the critical point enables us to make direct contact with the original situation to which no time-dependent field is applied, and thus to solve analytically the problem almost without any additional labor. We shall treat the cases of varying external field and varying temperature separately in the following.

### 2.1.1 Theory of field driving

In this case,  $H$  is varying with  $t$  linearly and  $\tau$  is a small constant. The theory can be rendered finite for  $d \leq 4$  by introducing the following  $Z$  factors (Amit & Martin-Mayor, 2005; Janssen,

1992; Zinn-Justin, 1996)

$$\begin{aligned} \varphi \rightarrow \varphi_0 &= Z_\varphi^{1/2} \varphi, & \tilde{\varphi} \rightarrow \tilde{\varphi}_0 &= Z_{\tilde{\varphi}}^{1/2} \tilde{\varphi}, & g \rightarrow g_0 &= N_d \mu^\epsilon Z_\varphi^{-2} Z_u u, \\ \lambda \rightarrow \lambda_0 &= (Z_\varphi / Z_{\tilde{\varphi}})^{1/2} Z_\lambda \lambda, & \tau \rightarrow \tau_0 &= Z_\varphi^{-1} Z_{\varphi^2} \tau + \tau_c, & H \rightarrow H_0 &= Z_\varphi^{-1/2} H, \end{aligned} \quad (19)$$

where  $\epsilon = 4 - d$ ,  $N_d = 2 / [(4\pi)^{d/2} \Gamma(d/2)]$  with  $\Gamma$  being the Euler Gamma function,  $\mu$  is an arbitrary momentum scale, and  $\tau_c$  the fluctuation shift of the critical point, which can be neglected as dimension regulations ('t Hooft & Veltman, 1972) are employed. However, we shall henceforth still use  $\tau$  to denote  $\tau - \tau_c$  that is proportional to  $T - T_c$ . Consequently, the critical point at  $\tau = 0$  and  $H = 0$  can be chosen to correspond to  $t = 0$  by a proper time translation. In Equation (19), the subscripts 0 indicate unrenormalized bare variables. By exploiting the fact that the bare quantities are independent of  $\mu$  and expanding the averaged order parameter in a Taylor's series in  $\tau$  and  $H$  at every definite time instant, namely

$$M(\tau, H) = \langle \varphi(\tau, H) \rangle = G_{10,0}(\tau, H) = \sum_{N, N'=1}^{\infty} \frac{1}{N! N'!} \lambda^{N+N'} \tau^{N'} H^N G_{1N, N'}(0, 0), \quad (20)$$

the Green function  $G_{1N, N'}$  is defined as

$$G_{1N, N'} = \frac{\delta^{1+N+N'} W[h, \tilde{h}, \tau]}{\delta h \delta \tilde{h}^N \delta \tau^{N'}} \quad (21)$$

the renormalization-group equation is thus,

$$\left( \mu \partial_\mu + \zeta \lambda \partial_\lambda + \beta \partial_u + \gamma_{\varphi^2} \tau \partial_\tau + \frac{1}{2} \gamma_H H \partial_H + \frac{1}{2} \gamma \right) M = 0, \quad (22)$$

with the Wilson's functions being defined as derivatives at fixed bare parameters,

$$\zeta(u) = \mu \partial_\mu \ln \lambda, \quad \gamma(u) = \mu \partial_\mu \ln Z_\varphi, \quad \gamma_{\varphi^2}(u) = \mu \partial_\mu \ln \tau, \quad \beta(u) = \mu \partial_\mu u, \quad (23)$$

where  $\partial_i$  indicates the partial derivative with respect to  $i$ . No new Wilson's function due to the driving has to be introduced.

At the fixed point

$$u = u^*, \quad \beta(u^*) = 0, \quad (24)$$

the solution of (22) is

$$M(\lambda, \tau, H, u, \mu) = \rho^{\gamma^*/2} M(\lambda \rho^{\epsilon^*}, \tau \rho^{\gamma_{\varphi^2}^*}, H \rho^{\gamma^*/2}, u^*, \mu \rho), \quad (25)$$

where  $\rho$  is a running (momentum) variable and starred quantities denote the corresponding values at the fixed point. On the other hand, from the naïve dimensions of various variables

$$\begin{aligned} |\mathbf{r}| &\propto \mu^{-1}, & \tau &\propto \mu^2, & u &\propto \mu^0, & \lambda t &\propto \mu^{-2}, \\ \varphi &\propto \mu^{(d-2)/2}, & H &\propto \mu^{(d+2)/2}, & \tilde{\varphi} &\propto \mu^{(d+2)/2}, \end{aligned} \quad (26)$$

one obtains a homogeneous form

$$M(\lambda, \tau, H, u, \mu) = \rho^{(d-2)/2} M(\lambda t \rho^2, \tau \rho^{-2}, H \rho^{-(d+2)/2}, u, \mu / \rho). \quad (27)$$

Applying (27) to the left-hand size of (25) leads to

$$M(t, \tau, H) = \rho^{\beta/\nu} M(t\rho^z, \tau\rho^{-1/\nu}, H\rho^{-\beta\delta/\nu}), \tag{28}$$

or in terms of the length variable  $b$ ,

$$M(t, \tau, H) = b^{-\beta/\nu} M(tb^{-z}, \tau b^{1/\nu}, Hb^{\beta\delta/\nu}), \tag{29}$$

with the critical exponents given by

$$\eta = \gamma^*, \quad \nu^{-1} = 2 - \gamma_{\phi^2}^*, \quad z = 2 + \zeta^*, \tag{30a}$$

$$\beta/\nu = (d - 2 + \eta)/2, \quad \delta = (d + 2 - \eta)/(d - 2 + \eta), \tag{30b}$$

where we have chosen  $\lambda$  as the time unit and used the same symbol to denote functions of different numbers with arguments.

As we perform the renormalization at the critical point and utilize the scheme of dimension regulations and minimal subtractions with  $\epsilon$  expansion ('t Hooft & Veltman, 1972), a scheme in which dynamics decouples from statics (De Dominicis & Peliti, 1978), all the  $Z$  factors can be chosen to be identical to the usual  $\phi^4$  model. As a result, all the static critical exponents and the dynamic critical exponent  $z$  determined from (30) are identical to those of the usual scalar Model A in the absence of the time-dependent field (Zhong, 2006; Zinn-Justin, 1996). Consequently, Equation (29) is just Equation (7) in the thermodynamic limit  $L^{-1} = 0$ .

Now the linearly varying field  $H = Rt$  can be complemented to Equation (29) (Zhong, 2006). (29) implies  $H$  and  $t$  transform as

$$H' = Hb^{\beta\delta/\nu}, \quad t' = tb^{-z}, \tag{31a}$$

respectively, where the primes indicate variables after rescaling. In the vicinity of  $T_c$ ,  $R$  should also scale upon renormalization. Suppose it transforms as

$$R' = Rb^{r_H}, \tag{31b}$$

since  $H' = R't'$ , one then obtains from Equation (31) a scaling law

$$r_H = z + \beta\delta/\nu, \tag{32}$$

which may be regarded as a definition of  $r_H$  that reflects the rescaling of  $R$  with that of  $H$  and  $t$  since  $R = H/t$ . Replacing  $t$  with  $R$  and setting  $b = R^{-1/r_H}$  in Equation (29), one finds

$$M(t, \tau, R) = R^{\beta/\nu r_H} m_{1H}(tR^{z/r_H}, \tau R^{-1/\nu r_H}), \tag{33}$$

or, in terms of some other pairs of the variables,

$$M(H, \tau, R) = R^{\beta/\nu r_H} m_{2H}(HR^{-\beta\delta/\nu r_H}, \tau R^{-1/\nu r_H}), \tag{34}$$

$$M(t, \tau, H) = H^{1/\delta} m_{3H}(tH^{\nu z/\beta\delta}, \tau H^{-1/\beta\delta}), \tag{35}$$

$$M(t, \tau, R) = t^{-\beta/\nu z} m_{4H}(Rt^{r_H/z}, \tau t^{1/\nu z}), \tag{36}$$

as only two out of the trio  $t$ ,  $R$ , and  $H$  are independent, where all  $m_{iH}$ s are scaling functions. Equations (33) to (36) are the finite-time scaling analogues of (3).

We have therefore justified the finite-time scaling. It is remarkable in this formulation that the critical exponents are naturally identical with the usual infinite time systems'. Also, no new independent exponent has to be introduced. In fact, since an expansion of the partition function in terms of a space-time dependent magnetic field generates correlation functions, the scaling properties of thermodynamic functions of a time-dependent magnetic field such as (33) and (34) follow naturally once the field is so small that the system still remains in the critical region.

### 2.1.2 Theory of temperature driving

In this case,  $H$  keeps zero, but  $T$  or  $\tau$  varies with time linearly. The scaling form can be derived following the procedure in the last section and can also be found in (Zhong, 2006). Here, we present an alternative semi-phenomenological derivation.

From Section 2.1.1, one can write directly the renormalization-group equation for the temperature driving as

$$\left( \mu \partial_\mu + \zeta t \partial_t + \beta \partial_u + \gamma_{\varphi^2} \tau \partial_\tau + \frac{1}{2} \gamma \right) M = 0, \quad (37)$$

whose solution at the fixed point, Equation (24), is

$$M(t, \tau, u, \mu) = \rho^{\gamma^*/2} M(t \rho^{\zeta^*}, \tau \rho^{\gamma_{\varphi^2}^*}, u^*, \mu \rho), \quad (38)$$

where we have directly used  $t$  in place of  $\lambda$ . However, as  $R$  is also a parameter, we may also write the renormalization-group equation as

$$\left\{ \mu \partial_\mu + \zeta t \partial_t + \beta \partial_u + \gamma_{\varphi^2} \tau \partial_\tau + r[u(\mu)] R \partial_R + \frac{1}{2} \gamma \right\} M = 0, \quad (39)$$

where we have assumed an additional Wilson function  $r(u)$  from a new renormalization factor associated with  $R$ . Its solution at the fixed point, Equation (24), is then

$$M(t, \tau, R, u, \mu) = \rho^{\gamma^*/2} M(t \rho^{\zeta^*}, \tau \rho^{\gamma_{\varphi^2}^*}, R \rho^{r^*}, u^*, \mu \rho), \quad (40)$$

where  $r^* = r(u^*)$ . Combining with the homogenous equation from dimensional analysis, Equation (27), results in

$$M(t, \tau, R, u, \mu) = \rho^{\beta/\nu} M(t \rho^z, \tau \rho^{-1/\nu}, R \rho^{-r_T}, u^*, \mu) \quad (41)$$

similar to Equation (28) with the exponents defined in Equation (30) and  $r_T = 4 - r^*$  since the naïve dimension of  $R$  is 4. However, because  $R = \tau/t$ , only the latter two variables are independent. As a consequence,

$$\partial_t f(t, \tau) = (\partial_t - \tau/t^2 \partial_R) w(t, \tau, R), \quad \partial_\tau f(t, \tau) = (\partial_\tau + 1/t \partial_R) w(t, \tau, R) \quad (42)$$

for two arbitrary derivable functions  $f$  and  $w$ . Substituting the derivative operators in Equation (42) into (37) and comparing with (39), one finds that

$$r(u) = \gamma_{\varphi^2}(u) - \zeta(u). \quad (43)$$

Therefore, at the fixed point, Equation (24), one has a scaling law

$$r_T = 4 - r^* = 4 - \gamma_{\varphi^2}(u^*) + \zeta(u^*) = z + 1/\nu \quad (44)$$

using Equation (30a). Equation (44) can of course be derived as in Section 2.1.1 from a length scaling version of Equation (41),

$$M(t, \tau, R) = b^{-\beta/\nu} M(tb^{-z}, \tau b^{1/\nu}, Rb^{r_T}), \quad (45)$$

which implies

$$t' = tb^{-z}, \quad \tau' = \tau b^{1/\nu}, \quad R' = Rb^{r_T}, \quad (46)$$

relating variables before and after a length rescaling of factor  $b$  and  $\tau = Rt$  (Zhong, 2006).

From Equation (45), finite-time scaling form for the temperature driving can be derived (Zhong, 2006) as

$$M(\tau, R) = R^{\beta/\nu r_T} m_{1T}(\tau R^{-1/\nu r_T}). \quad (47)$$

Similarly, finite-time scaling forms for the non-equilibrium susceptibility and specific heat that are defined respectively as the fluctuations of the order parameter and the energy  $E$  like their equilibrium counterparts are

$$\chi(T, R) = R^{-\gamma/\nu r_T} m_{2T}(\tau R^{-1/\nu r_T}), \quad (48)$$

$$C(T, R) = R^{-\alpha/\nu r_T} m_{3T}(\tau R^{-1/\nu r_T}). \quad (49)$$

## 2.2 Crossover

We analyze the crossover between the regime of finite-time scaling and that of equilibrium in this section.

### 2.2.1 Field driving

In the case of field driving, the finite-time scaling regime is defined by

$$\tau R^{-1/\nu r_H} \lesssim 1, \quad \tau H^{-1/\beta\delta} \lesssim 1, \quad \tau t^{1/\nu z} \lesssim 1. \quad (50)$$

Note that

$$|\tau|^{\nu r_H} = |\tau|^{\nu z} |\tau|^{\beta\delta} \propto \xi_{\infty}^{-z} M_{eq}^{\delta} \propto H_{eq}/t_{eq} \equiv R_{eq} \quad (51)$$

using Equations (5), (6), and (32), where  $H_{eq}$  is an equilibrium magnetic field corresponding to an equilibrium magnetization  $M_{eq}$  at  $\tau < 0$  and  $R_{eq}$  is an equilibrium rate. Equation (50) implies just

$$R \gtrsim R_{eq}, \quad H \gtrsim H_{eq}, \quad t \lesssim t_{eq}, \quad (52)$$

respectively. Accordingly, the finite-time scaling regime is characterized by an external rate and field that are larger than their corresponding intrinsic ones and an external time that is shorter than the intrinsic one and thus all these external scales become relevant similar to the case of finite-size scaling. As they all originate from the external driving, it is therefore reasonable and simpler to say that the finite-time scaling regime is characterized by an effective time scale  $R^{-1}$  shorter than the equilibrium correlation time  $t_{eq}$ . While in the reverse cases, for large  $\tau R^{-1/\nu r_H}$  or small  $R \ll R_{eq}$  for instance, the field varies so slowly that although it is changing, before it changes, the system has already equilibrated so that the

usual equilibrium scaling

$$M(\tau, H) = \tau^\beta f_4(H\tau^{-\beta\delta}) \tag{53}$$

emerges. Therefore, the scaling functions  $m_{iH}$  have a similar asymptotic behavior as,

$$m_i(x, y) \rightarrow \begin{cases} m_{5H}(x), & \text{for } y \rightarrow 0, \\ y^\beta f_4(xy^{-\beta\delta}), & \text{for } y \rightarrow \infty, \end{cases} \quad (i = 1, 2, 3, 4). \tag{54}$$

The crossover occurs when  $R \sim R_{eq}$  or  $H \sim H_{eq}$  or  $t \sim t_{eq}$ .

Although the finite-time scaling regime is defined by a large  $R$ , for too large  $R$  corresponding to a large  $H$  and short  $t$ , on the other hand, the system under such a driving is too far away from equilibrium and may enter another regime.

**2.2.2 Temperature driving**

In this case, there is only one scaled argument,  $\tau R^{-1/vr_T}$ , in the scaling functions. The finite-time scaling regime is thus defined by the smallness of this argument, viz.,  $|\tau|R^{-1/vr_T} \lesssim 1$ , or,

$$|\tau|^{-vr_T} R = |\tau|^{-vz} |\tau|^{-1} R \propto \xi_\infty^z R / |\tau| \propto t_{eq} R / |\tau| \gtrsim 1, \tag{55}$$

where Equations (5), (6), and (44) have been used. The last part of Equation (55) expresses clearly that the finite-time scaling regime is correctly defined by an effective time scale  $R^{-1}$  that is far shorter than the equilibrium correlation time  $t_{eq}$ . Accordingly, in close similarity to finite-size scaling, in this regime, the relevant scale is  $\tau R^{-1}/t_{eq} = R_{eq}/R$ . In the other extreme, the external time scale is so longer than  $t_{eq}$  that although the temperature is changing, before it changes, the system has already equilibrated in a way that the usual equilibrium behavior  $M \propto \tau^\beta$  emerges independent of  $R$ . Therefore, all the scaling functions  $m_{iT}$  behave asymptotically as

$$m_{iT}(x) \rightarrow \begin{cases} \text{constant} & \text{for } x \rightarrow 0, \\ x^\beta & \text{for } x \rightarrow \infty, \end{cases} \quad (i = 1, 2, 3). \tag{56}$$

The crossover occurs near  $\tau R^{-1} \sim t_{eq}$ .

**2.3 Corrections to scaling**

So far, we have only considered the relevant variables such as  $H$  and  $\tau$ . If there exist irrelevant variables, then there will be corrections to scaling induced by them (Wegner, 1972). We shall briefly discuss this issue in this section.

Assume that the leading irrelevant variable is  $Y$  and its corresponding exponent  $\omega > 0$ , Equation (45), for instance, is modified to

$$M(T, R, Y) = b^{-\beta/v} M(\tau b^{1/v}, R b^{r_T}, Y b^{-\omega}) \tag{57}$$

by neglecting the dependent variable  $t$ . Accordingly, Equation (47) becomes

$$M(T, R, Y) = R^{\beta/vr_T} m_{4T}(\tau R^{-1/vr_T}, Y R^{\omega/r_T}). \tag{58}$$

So, even at  $\tau = 0$ ,

$$M(T_c, R, Y) = R^{\beta/vr_T} m_{5T}(Y R^{\omega/r_T}). \tag{59}$$

It is the scaling function  $m_{5T}$  that induces the leading algebraic corrections to scaling (Wegner, 1972). Exactly at the critical point,  $R = 0$  and the corrections disappear; while near it, one can

expand  $m_{5T}(x)$  at  $x = 0$  as a series of  $x = YR^{\omega/r_T}$ ,

$$M(T_c, R, Y) = R^{\beta/vr_T} (A_0 + A_1 YR^{\omega/r_T} + A_2 YR^{2\omega/r_T} + \dots), \tag{60}$$

where  $A_i$  are constants.

From Equations (1) and (11), one can also write down the scaling forms with corrections as

$$M(T, L, Y) = L^{-\beta/v} f_5(\tau L^{1/v}, YL^{-\omega}), \tag{61}$$

$$M(T, t, Y) = t^{-\beta/vz} f_{5t}(\tau t^{1/vz}, Yt^{-\omega/z}) \tag{62}$$

for finite-size scaling and short-time critical dynamics, respectively. One finds therefore that the correction-to-scaling exponent decreases sequentially from finite-size scaling ( $\omega$ ), to short-time critical dynamics ( $\omega/z$ ), and to finite-time scaling ( $\omega/r$ ). This implies the corrections vary quite gently in the latter as compared to the other two cases and may thus be ignored without large errors in the first approximations in estimating critical properties.

In addition, if there is a marginal variable (Wegner, 1972), logarithmic corrections to scaling appears. Finite-time scaling forms in the presence of logarithmic corrections can also be derived. We leave this for future publications.

### 2.4 Combined finite-time and finite-size scalings

Up to now in our discussions of finite-time scaling, we have implicitly assumed that the system size is infinite, i.e., in the thermodynamic limit. We now take the finite-size effects into account.

In this case, the transformation law for the order parameter in temperature driving, for example, is

$$M(T, R, L) = b^{-\beta/v} M(\tau b^{1/v}, Rb^{r_T}, L^{-1}b), \tag{63}$$

from which the finite-time and finite-size scaling form

$$M(T, R, L) = R^{\beta/vr_T} m_{6T}(\tau R^{-1/vr_T}, L^{-1}R^{-1/r_T}) \tag{64}$$

follows. There are then several consequences that can be drawn.

First, the regime of finite-time scaling is further restricted to  $L^{-1}R^{-1/r_T} \ll 1$  or  $R^{-1/r_T} \ll L$  besides  $\tau R^{-1} < t_{eq}$  from Equation (55). So, for sufficiently large lattice sizes, the finite-size effects can be ignored. For not so large lattice sizes but still in the finite-time scaling regime, there are corrections from the finite size. Yet, as  $L^{-1}R^{-1/r_T} \ll 1$ , the corrections may still be small and be neglected.

Second, for  $L^{-1}R^{-1/r_T} \gg 1$  but  $\tau R^{-1/vr_T} = \tau L^{1/v} (L^{-1}R^{-1/r_T})^{1/v} \ll 1$  or  $L \ll \tau^{-v} \propto \xi_\infty$ , the system then crossovers to the finite-size scaling regime.

Third, if  $\tau R^{-1/vr_T} \gg 1$  besides  $L^{-1}R^{-1/r_T} \gg 1$ , equilibrium follows.

Combining the above three cases, one finds that the scaling function  $m_{6T}$  behaves asymptotically as

$$m_{6T}(x, y) \rightarrow \begin{cases} m_{1T}(x), & \text{for } y \rightarrow 0 \ \& \ x \rightarrow 0, \text{ finite time scaling} \\ y^{\beta/v} f_1(xy^{-1/v}), & \text{for } y \rightarrow \infty \ \& \ x \rightarrow 0, \text{ finite size scaling} \\ x^\beta, & \text{for } y \rightarrow \infty \ \& \ x \rightarrow \infty, \text{ equilibrium} \end{cases} \tag{65}$$

with the scaling functions  $m_{1T}$  and  $f_1$  defined above in Equations (47) and (3a), respectively.

The crossover from finite-time scaling to finite-size scaling regime occurs near  $R^{-1/r_\tau} \sim L$  for sufficiently small  $\tau$ . In fact, the former regime can be regarded as an effective finite-size scaling regime in which the driving-induced effective length scale  $R^{-1/r_\tau}$  dominates  $L$ . This can be generalized to the concept of driving simulations by which other scalings that one needs or wants or may be difficult to consider can be simulated with driving-induced effective scales.

### 3. Methods of finite-time scaling

Currently, there are mainly two catalogs of methods that have been developed to estimate both static and dynamic critical exponents as well as the critical temperature on the basis of finite-time scaling. They are respectively based on the field driving and the temperature driving. The main point underlying the classification is that in the field driving, one has two variables,  $\tau$  and  $H$ , at ones disposal, while in the temperature driving, only  $\tau$  is at hand. As a result, to obtain all the critical exponents, one has to resort to other methods like Monte Carlo renormalization group. Of course, methods that combine some or all of these are possible. For example, the field driving with an extended dynamic Monte Carlo renormalization-group method was first applied to the first-order phase transitions in the two-dimensional Ising model (Zhong, 2002). Also, combining finite-time scaling with finite-size scaling may be helpful.

#### 3.1 Field-driving method

This method is based on Equation (34) (Gong et al., 2010; Huang et al., 2010). In the finite-time scaling regime, the external time scale dominates and drives the system off equilibrium. Hysteresis then emerges even at  $T_c$ . In order to deal with the situation of two variables in Equation (34), we scan  $H$  back and forth with the same rate  $R$  to form a hysteresis loop and integrate over  $H$  to get its area  $A = \oint M dH$ . We then obtain from Equation (34) finite-time scaling forms of the coercivity  $H_c$  at  $M = 0$ ,  $A$ , and its derivative as,

$$H_c(\tau, R) = R^{n_H} m_{6H}(\tau R^{-1/vr_H}), \quad (66a)$$

$$A(\tau, R) = R^{n_{aH}} m_{7H}(\tau R^{-1/vr_H}), \quad (66b)$$

$$\partial A(\tau, R)/\partial \tau = R^{a_1} m_{8H}(\tau R^{-1/vr_H}), \quad (66c)$$

with

$$n_H = \beta\delta/vr_H, \quad n_{aH} = \beta(\delta + 1)/vr_H, \quad a_1 = \beta(\delta + 1)/vr_H - 1/vr_H. \quad (67)$$

At  $\tau = 0$ , exact power laws

$$H_c(0, R) \propto R^{n_H}, \quad A(0, R) \propto R^{n_{aH}}, \quad \partial A(0, R)/\partial \tau \propto R^{a_1} \quad (68)$$

follow, from which  $n_H$ ,  $n_{aH}$ , and  $a_1$  can be determined. The critical temperature can also be determined by finding the temperature at which minimum deviations from the power law behavior, Equation (68), occurs from studying Equations (66). Combining the exponents found with the hyperscaling law  $\beta(\delta + 1) = dv$  from Equation (4), one can calculate all the static and dynamic critical exponents from

$$\begin{aligned} \delta &= n_H/(n_{aH} - n_H), \quad \beta/v = d(n_{aH} - n_H)/n_{aH}, \quad z = d(1 - n_H)/n_{aH}, \quad r_H = d/n_{aH}, \\ \beta &= (n_{aH} - n_H)/(n_{aH} - a_1), \quad v = n_{aH}/d(n_{aH} - a_1). \end{aligned} \quad (69)$$

Note that in Equations (69) the first line requires only  $n_H$  and  $n_{aH}$ , while the last line needs  $a_1$ , which usually has a large statistical error due to numerical derivatives.

Of course, other observables such as  $\chi$  (Huang et al., 2010) may also be employed.

The hysteresis critical exponents  $n_H$  and  $n_{aH}$  (or the rate exponent  $r_H = d/n_{aH}$ ) have a particular meaning. Due to the scaling laws, Equations (4) and (30b), usually two critical exponents suffice to determine others for equilibrium critical phenomena. However, as  $\delta$  is directly related to  $\eta$  via Equation (30b), knowing these two can only produce ratios of exponents instead of individual exponents, because

$$(\gamma/\nu) = (\beta/\nu)(\delta - 1), \quad (70a)$$

$$(\gamma/\nu) = 2 - \eta, \quad (70b)$$

$$2(\beta/\nu) + (\gamma/\nu) = d. \quad (70c)$$

In fact, these two exponents can be used to characterize the so-called ‘weak’ universality class in which exponent ratios instead of exponents themselves are identical (Suzuki, 1974). If dynamics is taken into account, one then needs  $z$  besides those two because in the usual critical dynamics,  $z$  is independent of the static ones. However, owing to the new scaling law, Equation (32), in the finite-time scaling, they are related. One can easily find indeed that  $n_H$  and  $n_{aH}$  (or  $r_H$ ) suffice to determine  $\delta$ ,  $\eta$ , and  $z$ .

### 3.2 Temperature-driving method: Finite-time scaling with Monte Carlo renormalization group

This method is based on Equation (47) for a temperature sweep. However, it can at best give rise to the exponent ratios and  $T_c$ . In order to obtain more information, one way is to combine it with an extended dynamic Monte Carlo renormalization-group approach (Zhong, 2002). This approach may be regarded as a direct realization of Equation (46). It consists in matching correlation functions on different-sized lattices at different levels of renormalization to obtain renormalization-group eigenvalues and hence associated exponents. As a method to estimate the blocked variables, one resorts to a nearest-neighbor correlation function  $G_{nm}$  defined on a system of size  $Lb$  and assumes that after one block, it exactly matches that of a smaller system of size  $L$  without blocked, viz.,

$$G'_{nm,Lb}(T'_p, R') = G_{nm,L}(T_{ps}, R_s), \quad (71)$$

where  $T_p$  is the temperature at the peak of  $G_{nm}$  and  $s$  indicates quantities on the small lattice. In other words, one identifies the blocked variables with their unblocked counterparts on the small lattice. Consequently, one finds

$$r_T = \log(R_s/R) / \log b, \quad \nu = \log b / \log(\tau_{ps}/\tau_p) \quad (72)$$

from Equation (46) and hence  $z$  from Equation (44). Moreover, as the two systems whose  $G_{nm}$ s are compared have the same size, size effects are thus reduced.

Iterating this blocking procedure produces a series of exponents which should be invariant after a couple of blockings that iterate away the irrelevant variables if there is a fixed point controlling the scaling behavior, because the correlation functions will then track each other.

Furthermore, combining the first two equations of Equation (46) at  $T_p$ , one finds an invariant constant

$$a \equiv \left( \frac{\tau_p}{R^{1/\nu r_T}} \right)' = \frac{\tau_p}{R^{1/\nu r_T}} \propto \left( \frac{\tau_p R^{-1}}{t_{eq}} \right)^{1/\nu r_T} \quad (73)$$

under rescaling, which reflects again the similarity with finite-size scaling in which the ratio of the correlation length  $\xi_L(T_c)$  of a system of size  $L$  at  $T_c$  to  $L$ ,  $\xi_L(T_c)/L$ , is scale invariant (Amit & Martin-Mayor, 2005). Therefore,

$$T_p = T_c + aR^{1/\nu r_T}, \quad (74)$$

which offers both a method to estimate  $T_c$  and also a consistent check of the hysteresis exponents  $1/\nu r_T$  with that derived from Equation (72). Equation (74) is reasonable because at  $R = 0$ , or equilibrium, the correlation function ought to exhibit a peak at  $T_c$ . At a finite-time scale  $R^{-1}$ , there is an overshoot or hysteresis embodied in  $T_p$  due to the driving out of equilibrium.

The fitting and the Monte Carlo renormalization-group method then provide  $T_c$ ,  $\nu$ ,  $r_T$ , and  $z$ . Like the case of field driving, there is a scaling law, Equation (44), relating the static exponent  $\nu$  to dynamic ones  $r_T$  and  $z$ . As a result, verifying one out of the three then verifying the other two since there is a consistent check of the correctness of  $1/\nu r_T$ . In order to obtain other exponents, one can invoke the finite-time scaling forms of the order parameter, Equation (47), and other observables such as the non-equilibrium susceptibility, Equation (48), and specific heat, Equation (49), all of which can be measured during the course of heating without the need for independent setups. As the arguments of the scaling functions  $m_{iT}$  have been known, one can then just adjust one exponent in each case to collapse the curves of various  $R$ s or fit  $M$  or  $\chi$  or  $C$  at  $\tau = 0$  or at their respective peaks if available similar to Equation (68) to obtain directly the corresponding exponents.

As a lot of critical exponents can be obtained independently, one can then test the scaling laws, Equations (4) and (70). This is important for two reasons. First, the scaling laws may be broken in some cases (Fisher, 1986; Grinstein, 1976) in which there is a dangerous irrelevant variable (Wegner, 1972). Second, if valid, they give strong evidences for the asymptotic nature of the exponents obtained which is not easily obtainable in disordered systems, because asymptotic exponents ought to satisfy scaling laws if they are valid.

## 4. Summary of results from finite-time scaling

We summarize briefly in this section the main results that have been obtained using the methods of finite-time scaling presented in Section 3. We again present them according to the classification we adhered to in this chapter.

### 4.1 Results obtained using field-driving method

The field-driving method has been applied successfully to the two-dimensional and three-dimensional Ising models (Gong et al., 2010) and the two-dimensional three- and four-state Potts models (Huang et al., 2010).

#### 4.1.1 Simulation details

For the Ising model (Ising, 1925) in the presence of an external field, there is an inversion symmetry, which also reflects in the hysteresis loops. The coercive field can then be simply

$d$	$n_{qH}$	$n_H$	$a_1$	$r_H$	$\delta$	$z$	$\beta$	$\nu$
2	0.4965(18)	0.4653(19)	0.244(7)	4.028(15)	14.9(1.3)	2.154(11)	0.124(10)	0.983(28)
	0.49487(15) <sup>†</sup>	0.46394(14) <sup>†</sup>	0.24743(8) <sup>†</sup>	4.0415(12) <sup>†</sup>	15 <sup>‡</sup>	2.1665(12) <sup>§</sup>	1/8 <sup>‡</sup>	1 <sup>‡</sup>
3	0.6650(28)	0.5466(22)	0.314(7)	4.511(19)	4.62(15)	2.045(13)	0.337(11)	0.632(13)
	0.6647(6) <sup>†</sup>	0.5499(5) <sup>†</sup>	0.3131(6) <sup>†</sup>	4.513(4) <sup>†</sup>	4.789(2) <sup>b</sup>	2.031(3) <sup>‡</sup>	0.3265(3) <sup>b</sup>	0.6301(4) <sup>b</sup>

Table 1. Measured and derived exponents of the Ising model. <sup>†</sup>: calculated from other exponents in the same row (Zhong, 2006); <sup>‡</sup>: exact results; <sup>§</sup>: from (Nightingale & Blöte, 1996) (Nightingale & Blöte, 2000); <sup>b</sup>: from (Pelissetto & Vicari, 2002); and <sup>‡</sup>: average of the values estimated by (Grassberger, 1995) and (Kikuchi & Ito, 1993).

defined as in Section 3.1. In the case of the  $q$ -state Potts model (Potts, 1952), we apply an external field along one Potts state. No inversion symmetry in the hysteresis loops exist any more. We then employ the peak of a non-equilibrium susceptibility as a definition of the coercivity. Periodic boundary conditions are applied throughout. We choose several temperatures around the critical point of a model and a series of rates at each temperature. There are several considerations for the rates chosen. First, they cannot be too large to avoid far away from equilibrium, which may be qualified when the equilibrium equation of state away from the transition region is followed. Second, they cannot be too small to avoid leaving the finite-time scaling regime for the equilibrium or finite-size scaling regime. Third, they should be as small as possible in order to reduce errors from the relatively large values of an observable at large rates as compared to the small ones. Then, for each chosen temperature and rate, we start a Monte Carlo simulation at an ordered state with a sufficiently large external field that is larger than the closure field of the hysteresis loops at that rate to ensure closure of the loops and that has otherwise been checked to have no effect on the results. Several Monte Carlo steps suffice to equilibrate the system as it is far away from its critical point. After equilibrium, the field is swept with the rate back and forth for 100 times, say, to obtain 100 hysteresis loops. Averaged values of  $A$  and  $M$  can then be obtained. The method presented in Section 3.1 then yields  $T_c$  and exponents.

#### 4.1.2 Results

The  $T_c$ s for both models determined from the minimum deviations of  $A$  from pure power laws agree well with their respective exact results. Knowing  $T_c$  can then give rise to the critical exponents, which are given in Tables 1 (Gong et al., 2010) and 2 (Huang et al., 2010). One sees that the present results agree reasonably well with the exact ones, showing the effectiveness of the method. Note that besides the early real-space renormalization-group studies (see (Wu, 1982) for a review), the present is the one that directly applies an external field to estimate  $\delta$ , though it can also be estimated by other critical exponents through scaling laws. Slightly overlapped within the statistical errors as they are, our  $\delta$ s for  $q = 3$  and 4 still support their respective conjectured values. The dynamic critical exponent  $z$ s obtained for both the  $q = 3$  and  $q = 4$  Potts model agree well with previous Monte Carlo simulation results (de Alcantara Bonfim, 1987; Tang & Landau, 1987). Along with  $z$  in Table 1 of the two-dimensional Ising model that is equivalent to the  $q = 2$  Potts model, they appear to confirm the dynamic weak universality (de Alcantara Bonfim, 1987; Tang & Landau, 1987) according to which the Potts model with  $q = 2, q = 3$ , and  $q = 4$  all share the same  $z$ . However, they are apparently distinct from the short-time dynamic results of  $z = 2.29$  for  $q = 4$  (da Silva et al., 2002; Fernandes et al., 2006) but  $z = 2.19$  for  $q = 3$  (da Silva et al., 2002; Okano et al., 1997; Zhang et al., 1999). Thus, further studies are still needed here.

$q$	$n_{aH}$	$n_H$	$a_1$	$r_H$	$\delta$	$\beta/\nu$	$\beta$	$\nu$	$z$
3	0.4969(12)	0.4624(9)	0.200(10)	4.025(10)	13.4(6)	0.139(6)	0.116(5)	0.838(3)	2.164(7)
	0.4962(8) <sup>‡</sup>	0.4631(8) <sup>‡</sup>	0.1985(10) <sup>‡</sup>	4.031(7) <sup>‡</sup>	14 <sup>†</sup>	2/15 <sup>†</sup>	1/9 <sup>†</sup>	5/6 <sup>†</sup>	
4	0.4954(13)	0.4645(21)	0.154(12)	4.039(11)	15.1(1.3)	0.124(10)	0.0900(7)	0.726(3)	2.163(10)
	0.4953(12) <sup>‡</sup>	0.4643(11) <sup>‡</sup>	0.1238(15) <sup>‡</sup>	4.038(10) <sup>‡</sup>	15 <sup>†</sup>	1/8 <sup>†</sup>	1/12 <sup>†</sup>	2/3 <sup>†</sup>	

Table 2. Measured and derived exponents of the two-dimensional Potts model. <sup>†</sup>: conjectured values (Wu, 1982); <sup>‡</sup>: calculated from the conjectured values and the measured  $z$ s.

## 4.2 Results obtained using temperature-driving method

This method has been applied successfully to pure systems including the two-dimensional Ising model (Zhong & Xu, 2005) and the three-state Potts model (Fan & Zhong, 2007), and disordered systems including a two-dimensional random-bond Potts model with the state number  $q = 5$  and  $q = 8$  (Fan & Zhong, 2009), a three-dimensional random-bond Ising model (Xiong et al., 2010a), and a three-state random-bond Potts model (Xiong et al., 2010b).

### 4.2.1 Simulation details

In this case, in addition to the general considerations and boundary conditions given in Section 4.1.1, a pair of lattice sizes, for example, 128 and 64 in three dimensions has to be used owing to the two-lattice matching in renormalization. For a given  $R$  and, in the case of disordered systems, a disorder strength and a sample of its fixed realization, we start a Monte Carlo simulation from a completely ordered state at an initial temperature that is chosen to be so far away from  $T_p$  that it has been checked to have no effect on the results. After one time unit consisting of a sequential sampling of all the spins with the usual Metropolis algorithm (Metropolis et al., 1953),  $T$  is increased by  $R$ . The system then evolves with time until a disordered state is reached. At each time in the course of heating, we calculate a sample of  $M$ ,  $G_{nn}$ , and  $E$  and perform sequentially blockings on the configurations from which the renormalized  $G_{nn}$  is computed by means of a majority rule with  $b = 2$ . Ties are broken by a random selection among the tied states. Each quantity is then averaged at each time step over different samples.

### 4.2.2 Results

So far, the critical temperatures obtained agree quite well with either exact results (Fan & Zhong, 2007; 2009; Zhong & Xu, 2005) or existing estimates (Fan & Zhong, 2009; Xiong et al., 2010b) or an approximate theory (Xiong et al., 2010a;b). The critical exponents obtained for the pure two-dimensional Ising model,  $\nu = 0.97(8)$ ,  $z = 2.15(13)$ , and  $\beta = 0.12(1)$  for only 8 to 40 samples, and for the pure two-dimensional three-state Potts model,  $\nu = 0.816(27)$ ,  $z = 2.171(62)$  and  $\beta = 0.108(4)$  for more than 200 samples and hence smaller statistical errors, agree well with the corresponding values listed in Tables 1 and 2. A positive  $\alpha = 0.368(54)$  calculated from the hyperscaling law, Equation (4a), has also been testified by a scaling collapse of the specific heat curves for the latter model (Fan & Zhong, 2007).

Disorder is ubiquitous; its effects on critical behavior are thus important. This has become clear from Harris criterion (Cardy, 1996; Harris, 1974), namely, uncorrelated quenched randomness coupled to local energy density is irrelevant and the universality class of the pure system persists when its specific heat critical exponent  $\alpha < 0$ , while such randomness will lead to a new universality class controlled by a new 'random' fixed point when  $\alpha > 0$ . To identify the asymptotic critical exponents that characterize the random fixed point in the latter case is, however, not a simple task, as disorder-dependent critical exponents are frequently obtained, possibly reflecting the competition of different fixed points (Berche & Chatelain,

$r_0$	$q$	$r_T$	$\nu$	$1/\nu r_T$	$z$	$\beta$	$\beta/\nu$	$\alpha = 2 - d\nu$
3	8	4.39(12)	0.757(43)	0.302(16)	3.07(12)	0.130(15)	0.172(10)	0.49(9)
	5	4.15(8)	0.818(43)	0.295(13)	2.93(7)	0.128(10)	0.157(4)	0.36(9)
10	8	5.43(13)	1.021(35)	0.181(7)	4.45(13)	0.170(20)	0.167(14)	-0.042(70)
	5	5.12(7)	1.027(23)	0.190(5)	4.15(7)	0.161(20)	0.157(16)	-0.054(46)
15	8	5.75(15)	1.098(27)	0.158(6)	4.84(15)	0.173(20)	0.158(14)	-0.196(54)
	5	5.44(10)	1.112(34)	0.167(5)	4.54(9)	0.164(20)	0.148(14)	-0.22(7)
20	8	5.88(17)	1.180(35)	0.144(5)	5.03(17)	0.176(20)	0.149(13)	-0.36(7)
	5	5.62(14)	1.182(34)	0.151(5)	4.77(14)	0.170(20)	0.144(13)	-0.36(7)

Table 3. Critical exponents of the two-dimensional random-bond Potts model.

$r_0$	$r_T$	$\nu$	$1/\nu r_T$	$z$	$\beta$	$\alpha$	$\gamma$	$\beta/\nu$	$\gamma/\nu$
2	3.60(6)	0.651(18)	0.427(7)	2.061(32)	0.374(6)	-0.035(16)	1.389(18)	0.575(18)	2.13(7)
4	3.57(5)	0.682(18)	0.410(7)	2.108(35)	0.349(6)	-0.046(17)	1.330(22)	0.512(16)	1.95(7)
5	3.57(5)	0.689(18)	0.407(7)	2.119(37)	0.343(6)	-0.052(17)	1.333(22)	0.498(16)	1.93(7)
10	3.48(5)	0.765(19)	0.376(6)	2.175(35)	0.354(6)	-0.130(29)	1.420(32)	0.463(17)	1.86(5)

Table 4. Critical exponents of the three-dimensional random-bond Ising model.

$r_0$	$r_T$	$\nu$	$1/\nu r_T$	$z$	$\beta$	$\alpha$	$\gamma$	$\beta/\nu$	$\gamma/\nu$
2.5	4.28(3)	0.518(11)	0.451(7)	2.38(3)	0.206(8)	0.54(2)	1.16(2)	0.40(2)	2.24(6)
5	4.30(4)	0.540(9)	0.431(6)	2.44(3)	0.25(2)	0.48(3)	1.15(2)	0.46(4)	2.13(5)
7.5	4.32(5)	0.542(10)	0.426(6)	2.47(4)	0.29(3)	0.40(3)	1.15(2)	0.54(6)	2.12(5)
10	4.31(5)	0.554(9)	0.419(6)	2.51(4)	0.30(3)	0.36(3)	1.15(3)	0.54(6)	2.08(7)
15	4.34(6)	0.566(15)	0.408(7)	2.57(3)	0.31(3)	0.29(4)	1.15(5)	0.55(6)	2.03(10)
20	4.34(7)	0.569(16)	0.406(9)	2.58(5)	0.32(3)	0.28(4)	1.15(5)	0.56(6)	2.02(12)
30	4.21(11)	0.673(25)	0.353(9)	2.72(8)	0.34(4)	0.27(4)	1.25(5)	0.51(6)	1.86(10)

Table 5. Critical exponents of the three-dimensional three-state random-bond Potts model.

2004; Folk et al., 2003). We have studied three random-bond models in which the single pure bonds can randomly select between a weak bond  $K$  and a strong one  $r_0K$  with equal probability with  $r_0$  characterizing the disorder strength. The three-dimensional random-bond Ising model in its pure version has a continuous transition with a positive  $\alpha$ , while the two- and three-dimensional Potts models studied have first-order phase transitions in their pure version and disorders make them continuous (Aizenman & Wehr, 1989; Cardy & Jacobsen, 1997; Hui & Berker, 1989). All the three models will thus exhibit new universality classes in principle.

The exponents obtained using the method detailed in Section 3.2 of the three models are listed in Tables 3 to 5. They have been checked to be independent of the lattice sizes used for some disorder strengths in all the models. Generally speaking, our exponents agree quite well with existing results except  $\nu$  and  $\alpha$  of the latter model (Xiong et al., 2010b). Details can be found in the original papers and we shall not discuss them here to save space. Rather, we shall only focus on those special aspects.

For the two-dimensional random-bond Potts model, its exponents in Table 3 exhibit two distinct regimes with  $\alpha$  showing opposite signs, which, as indicated, is calculated by the scaling law, Equation (4a), and has been checked by scaling collapses. A positive  $\alpha$  means  $\nu < 1$  as seen, which violates the bound

$$\nu \geq 2/d \tag{75}$$

suggested to be satisfied for disordered systems (Chayes et al., 1986). This violation was also found in (Cardy & Jacobsen, 1997) but was later argued to be due to the insufficient disorder

strength because for  $q \gtrsim 2$  it was found that the bound (75) was satisfied for stronger disorders in agreement with our results, suggesting the violation was a result of crossover from the pure fixed point to the random fixed point (Jacobsen, 2000). However, in our case,  $q > 4$  and the pure model exhibits a discontinuous rather than a continuous transition. Although the hyperscaling law, Equation (4a), has been testified in this case and hence the activated dynamics proposed originally for the random-field Ising model (Fisher, 1986) is excluded (Deroulers & Young, 2002), as no further exponents have been obtained, one cannot draw definite conclusions about this regime. For  $q = 8$ ,  $r_0 \sim 10$  was found to be close to the random fixed point (Cardy & Jacobsen, 1997) and our  $\beta/\nu$  of  $r_0 = 10$  appears to be a little larger than 0.142(1) (Cardy & Jacobsen, 1997), 0.153(3) (Chatelain & Berche, 1998), and 0.153(1) (Jacobsen & Picco, 2000). Yet,  $r_0 = 8$  to 20 was found to locate the random fixed point with  $\beta/\nu = 1.50$  to 1.55 (Picco, 1998). If we averaged the three values within this range, we would get  $\beta/\nu = 1.58(8)$  that would agree quite well with those quoted values and also with 0.157(2)/0.156(11) of short-time critical dynamics (Yin et al., 2004). The same average yields  $\nu = 1.100(19)$  which appears again a little larger than about 1.02 (Cardy & Jacobsen, 1997; Chatelain & Berche, 1998). However,  $\nu$  was found to increase slightly with  $q$  with the  $q = 3$  value of 1.02(2) (Jacobsen, 2000). So, a slightly larger  $\nu$  for larger  $q$  may be still possible. We may also consider averages over  $r_0 = 15$  and 20 whose exponents appear closer in value. Anyway, the true critical exponents for the random fixed point in this model still need further studies.

The three-dimensional disordered Ising model as a paradigm of a positive  $\alpha$  in the pure case has attracted much interest (Folk et al., 2003) and its renormalization-group theory has reached a level of up to six loops (Pelissetto & Vicari, 2000). However, problems still exist concerning for example its true critical exponents (Xiong et al., 2010a). As a lot of exponents can be estimated, we are able to test the scaling laws as shown in Table 6. One sees that in the middle range of disorders, the exponents satisfy the three scaling laws tested and vary little. The averaged exponents within this range are thus regarded as the asymptotic critical exponents of the random fixed point. They agree well with results of other types of disorders and of the renormalization-group theory. These results thus lead to several conclusions. First, for the random fixed point, we have proved the validity of the scaling laws, which was invoked previously to reckon the correctness of the obtained exponents (Pelissetto & Vicari, 2000). Second, they help to unify the exponents. For example, our dynamic critical exponent of  $z = 2.114(51)$  supports a lower value found by renormalization-group analyzes, experiments, and some Monte Carlo simulations rather than the larger values of  $z \approx 2.6$  (Parisi et al., 1999; Schehr & Paul, 2005) and  $z \approx 2.35$  (Calabrese et al., 2008; Hasenbusch et al., 2007). Third, they corroborate the universality of the random fixed point with respect to the form of disorders. Fourth, they show that corrections to scaling can indeed be ignored in estimating exponents in finite-time scaling. Fifth, they also demonstrate the effectiveness of finite-time scaling in probing both static and dynamic critical behavior. The exponents at  $r_0 = 2$  and  $r_0 = 10$  do not satisfy all scaling laws and may thus be crossover exponents that reflect crossover from the random fixed point to the pure and to the percolation fixed point, respectively. Conversely, validating of a single or even two scaling laws may not be invoked as an indication of the asymptotic nature of the obtained exponents.

The three-dimensional random-bond Potts model shows gross features that are similar to the random-bond Ising model. In particular, one finds from Table 7 that the first two scaling laws are satisfied within the errors for  $r_0 \simeq 10$  to 20, and almost satisfied for  $r_0 = 7.5$ , but not for the other disorder strengths, while the third law can be considered as satisfied for all  $r_0$

$r_0$	2	4	5	10	Exact value
$\alpha + d\nu$	1.92(6)	2.00(6)	2.01(6)	2.17(6)	2
$\alpha + 2\beta + \gamma$	2.10(3)	1.98(3)	1.97(3)	2.00(5)	2
$2\beta/\nu + \gamma/\nu$	3.28(10)	2.97(9)	2.93(8)	2.79(8)	3

Table 6. Test of scaling laws for the three-dimensional random-bond Ising model.

$r_0$	2.5	5	7.5	10	15	20	30	Exact value
$\alpha + d\nu$	2.09(4)	2.10(4)	2.03(4)	2.01(5)	1.97(6)	1.99(6)	2.30(9)	2
$\alpha + 2\beta + \gamma$	2.10(3)	2.13(5)	2.13(7)	2.11(7)	2.06(9)	2.07(9)	2.20(10)	2
$2\beta/\nu + \gamma/\nu$	3.04(8)	3.05(9)	3.20(13)	3.16(14)	3.13(16)	3.14(16)	2.88(18)	3

Table 7. Test of scaling laws for the three-dimensional random-bond Potts model.

studied. Therefore, conclusions similar to the Ising model can also be drawn. For example, the exponents within  $r_0 \simeq 10$  to 20 are asymptotic and controlled by the random-fixed point in the model while those outside are only crossover. Corrections to scaling can again be ignored, etc.

However, there is an important difference.  $\alpha$  for the random fixed point is positive and thus  $\nu$  violates the bound (75) in contrary to recent numerical studies (Ballesteros et al., 2000; Chatelain et al., 2001; 2005; Mercaldo et al., 2005; 2006; Murtazaev et al., 2007; 2008; Yin et al., 2005; 2006) and a renormalization-group analysis (Aharony et al., 1998). In the two-dimensional random-bond Potts model, a positive  $\alpha$  has also been found as pointed out above. However, for large disorder strengths,  $\alpha$  becomes negative and  $\nu$  satisfies the bound. In the present case,  $\alpha$  is still a large positive number even for  $r_0 = 30$ , whose  $\nu = 0.673(25)$  is on the verge of  $2/d$  albeit with a large error. A hint for a positive  $\alpha$  has also been found but with  $\nu > 2/d$  in a three-dimensional random-bond Potts model in the large- $q$  limit (Mercaldo et al., 2005; 2006). Yet, the author claimed that the asymptotic region for the specific heat was far from the possibilities of present-day numerical calculations (Mercaldo et al., 2005; 2006). Negative  $\alpha$ s have been obtained on a small range of lattice sizes ( $L = 20 - 44$ ) using finite-size scaling but without considering corrections to scaling (as the exponent is several times bigger than that of finite-time scaling, Section 2.3) and without showing scaling collapse for the three-dimensional three-state Potts model with site dilutions (Murtazaev et al., 2007; 2008). This was obtained by fitting the peaks of the specific heat to

$$C = c_1 - c_2 L^{\alpha/\nu} \tag{76}$$

for a negative  $\alpha$ , where  $c_1$  and  $c_2$  are positive constants. We have found that in some ranges of rates, a fit to Equation (76) with  $L$  replaced by  $R^{-1/r}$  according to Section 2.4 does give a negative  $\alpha$ , but the  $C$  curves collapse badly even we adjust  $\alpha$  in the negative region (Xiong et al., 2010b). On the contrary, in the case of the random-bond Ising model, fits to such a form indeed lead to those negative  $\alpha$ s listed in Table 4, which collapse the specific curves well. In contrast, fits to the positive  $\alpha$  forms can also yield positive  $\alpha$ s but then the specific curves collapse badly (Xiong et al., 2010a). In addition, in the two-dimensional three-state pure Potts model, we have essentially applied the same methods to correctly identify its positive  $\alpha$  as mentioned in the first paragraph in this section. Moreover, all exponent ratios agree well with existing ones. Furthermore,  $z$  agrees well with that from short-time critical dynamics (Yin et al., 2005). This single exponent then lends support to our  $\nu$  and through Equation (4a)  $\alpha$  as pointed out in Section 3.2. All these therefore strongly support our positive  $\alpha$  and  $\nu < 2/d$ .

In fact, for a dirty system, it has been known that its stability is not directly related to its  $\alpha$ , which may thus assume positive values (Andelman & Berker, 1984; Kinzel & Domany, 1981), though the opposite is true for a pure system according to the Harris criterion (Harris, 1974). Moreover, it has also been pointed out that for systems in which self-averaging breaks down (Aharony & Harris, 1996; Wiseman & Domany, 1995; 1998), the  $\nu$  that is found by finite-size scaling and was proved to satisfied the bound (75) (Chayes et al., 1986) may be different from the intrinsic  $\nu$  that might escape it, since the former is found to be only a result of the grand canonical ensemble average used (Pazmandi et al., 1997), though a renormalization-group analysis shows that the average procedure is irrelevant (Aharony et al., 1998). If this is true, finite-time scaling will be superior.

## 5. Conclusion

We have reviewed in this Chapter the idea, the theory, and the methods of finite-time scaling and the results of their applications to the continuous phase transitions in both pure and disordered two- and three-dimensional Ising and Potts models. Both field driving and temperature driving have been considered. Both static and dynamic critical exponents as well as the critical points can all be estimated. As a lot of exponents can be determined independently, scaling laws can be tested, which is a valuable information for reckoning the asymptotic nature of the exponents. So far, most results obtained agree quite well with those from other sources and those disagreed appear quite possibly true. If the latter is shown to be correct, finite-time scaling will be superior to finite-size scaling. Even if it were finally shown to be wrong, the former results still have already demonstrated its effectiveness; and the lessons gained would certainly push it forward. We conclude that the idea behind finite-time scaling is physically so simple and in so close analogue to that of finite-size scaling that it should at least be a useful concept in statistical physics.

To end the review, we remarks on some other advantages and disadvantages of the finite-time scaling. It is a nonequilibrium approach that drives a system out of equilibrium. As a consequence, hysteresis ensues even at  $T_c$ . It is distinct from usual approaches in that it manipulates the dynamics of a system by an external driving field or temperature. This enables it to avoid critical slowing down (Gong et al., 2010; Huang et al., 2010). As has been pointed out, the correction-to-scaling exponent is rather small compared to finite-size scaling and short-time critical dynamics. This has two sides. On the one hand, it makes estimation of exponents rather simple since the corrections appear negligible. Moreover, the error bars of the exponents so estimated are also on a par with other usual methods. On the other hand, if one wants to make more precise estimations including the correction-to-scaling exponent, large ranges of time scales appear necessary. We have so far concentrated on the local dynamics as realized in the Monte Carlo simulations of single-site Metropolis algorithm and their equivalent Langevin dynamics, the idea of finite-time scaling, however, should be applicable to other dynamics as well. Finally, we would point out that the method of linear driving may probably be the simplest but most general approach to finite-time scaling and should also be amenable to experiments (Gong et al., 2010). It may also be generalized to a concept of driving simulations (Section 2.4) that apply the linear driving to simulate other effects like system sizes near criticality.

Future studies may include applying and testing finite-time scaling in other systems including quantum ones, exploiting the combined scalings of both finite times and finite sizes, developing approaches to improve the precision of the present methods, and applying the

scaling experimentally to study critical phenomena, etc. Another area that finite-time scaling is helpful is the scaling behavior in first-order phase transitions by driving (Zhang et al., 1995). In fact, the renormalization-group theory for the linear driving developed first in this area (Zhong & Chen, 2005). Moreover, the linear driving may possibly be crucial here (Fan & Zhong, 2010).

## 6. Acknowledgement

I am grateful to my professor, Professor Jinxiu Zhang. My use of the linear driving method originated from him. I am also indebted to my students, Zhifang Xu, Shuangli Fan, Shurong Gong, Xianzhi Huang, and Wanjie Xiong, whose practices and discussions drove me to develop the theory and methods presented here. ZX initiated simulations in the continuous transition of the Ising model. SF first considered the specific heat and WX included the susceptibility. SG first extended the renormalization-group theory to the off-critical situation. In addition, most of the numerical results presented here were obtained by them. Another student, Shuai Yin, always took part actively in our group meetings and his helpful discussions in completing the other derivation of the scaling law, Equation (44), and others are also appreciated. This work was supported by the National Natural Science Foundation of China (Grant Nos. 10374118 and 10625420), the FANEDD and EYTP of MOE, China, and NSF of Guangdong Province, China (Grant No. 011140).

## 7. References

- Aharony A. & Harris, A. B. (1996). Absence of self-averaging and universal fluctuations in random systems near critical points. *Phys. Rev. Letts.*, Vol. 77, 3700-3703.
- Aharony, A.; Harris, A. B. & Wiseman, S. (1998). Critical disordered systems with constraints and the inequality  $\nu > 2/d$ . *Phys. Rev. Lett.*, Vol. 81, 252-255.
- Aizenman, M. & Wehr, J. (1989). Rounding of first-order phase transitions in systems with quenched disorder. *Phys. Rev. Lett.*, Vol. 62, 2503-2506.
- Amit, D. J.; Martin-Mayor, V. (2005). *Field Theory, the Renormalization Group, and Critical Phenomena*, 3rd edition, World Scientific, Singapore.
- Andelman D. & Berker, A. N. (1984). Scale-invariant quenched disorder and its stability criterion at random critical points. *Phys. Rev. B*, Vol. 29, 2630-2635.
- Ballesteros, H. G.; Fernández, L. A.; Martín-Mayor, V.; Muñoz Sodupe, A.; Parisi, G. & Ruiz-Lorenzo, J. J. (2000). Critical behavior in the site-diluted three-dimensional three-state Potts model. *Phys. Rev. B*, Vol. 61, 3215-3218.
- For a review, see Berche, B. & Chatelain, C. (2004). Phase transitions in two-dimensional random Potts models. In: *Order, Disorder, and Criticality*, Holovatch, Yu. (Ed.), 147-199, World Scientific, Singapore.
- Binder, K. & Heermann, D. (1988). *Monte Carlo Simulations in Statistical Physics*, Springer, Berlin.
- Brézin, E. (1982). An investigation of finite size scaling. *J. de Phys.*, Vol. 43, 15-22.
- Brézin, E. & Zinn-Justin, J. (1985). Finite size effects in phase transitions. *Nucl. Phys. B*, Vol. 257, 867-893.
- Calabrese, P.; Pelissetto, A. & Vicari, E. (2008). Static and dynamic structure factors in three-dimensional randomly diluted Ising models. *Phys. Rev. E*, Vol. 77, 021126.
- Cardy, J. (1988). *Finite Size Scaling*, (Ed.), North-Holland, Amsterdam.
- Cardy, J. (1996). *Scaling and Renormalization in Statistical Physics*, Cambridge, Cambridge.

- Cardy, J. & Jacobsen, J. L. (1997). Critical behavior of random-bond Potts models. *Phys. Rev. Lett.*, Vol. 79, 4063-4066.
- Chayes, J. T.; Chayes, L.; Fisher, D. S. & Spencer, T. (1986). Finite-size scaling and correlation lengths for disordered systems. *Phys. Rev. Lett.*, Vol. 57, 2999-3002.
- Chatelain C. & Berche, B. (1998). Finite-size scaling study of the surface and bulk critical behavior in the random-bond eight-state Potts model. *Phys. Rev. Lett.*, Vol. 80, 1670-1673.
- Chatelain, C.; Berche, B.; Janke, W. & Berche, P. E. (2001). Softening of first-order transition in three-dimensions by quenched disorder. *Phys. Rev. E*, Vol. 64, 036120.
- Chatelain, C.; Berche, B.; Janke, W. & Berche, P. E. (2005). Monte Carlo study of phase transitions in the bond-diluted 3D 4-state Potts model. *Nucl. Phys. B*, Vol. 719 [FS], 275-311.
- da Silva, R.; Alves, N. A. & Drugowich de Felício, J. R. (2002). Mixed initial conditions to estimate the dynamic critical exponent in short-time Monte Carlo simulation. *Phys. Lett. A*, Vol. 298, 325-329.
- de Alcantara Bonfim, O. F. (1987). Critical dynamics of the  $q$ -state Potts model in two dimensions. *Europhys. Lett.*, Vol. 4, 373-376.
- De Dominicis, C. & Peliti, L. (1978). Field-theory renormalization and critical dynamics above  $T_c$ : Helium, antiferromagnets, and liquid-gas systems. *Phys. Rev. B*, Vol. 18, 353-376.
- Deroulers C. & Young, A. P. (2002). Critical behavior and lack of self-averaging in the dynamics of the random Potts model in two dimensions. *Phys. Rev. B*, Vol. 66, 014438.
- Fan, S. & Zhong, F. (2007). Determination of the dynamic and static critical exponents of the two-dimensional three-state Potts model using linearly varying temperature. *Phys. Rev. E*, Vol. 76, 041141.
- Fan, S. & Zhong, F. (2009). Critical dynamics of the two-dimensional random-bond Potts model with nonequilibrium Monte Carlo simulations. *Phys. Rev. E*, Vol. 79, 011122.
- Fan, S. & Zhong, F. (2010). Evidences for the instability fixed points of first-order phase transitions. submitted.
- Fernandes, H. A.; Arashiro, E.; Drugowich de Felício, J. R. & Caparica, A. A. (2006). An alternative order parameter for the 4-state potts model. *Physica A*, Vol. 366, 255-264.
- Ferrell, R. A.; Menyhárd, N.; Schmidt, H.; Schwabl, F. & Szépfalussy, P. (1967). Dispersion in second sound and anomalous heat conduction at the lambda point of liquid helium. *Phys. Rev. Lett.*, Vol. 18, 891-894.
- Fisher, M. E. (1967). Theory of condensation and critical point. *Physics*, Vol. 3, 255-283.
- Fisher, M. E. & Ferdinand, A. E. (1967). Interfacial, boundary, and size effects at critical points. *Phys. Rev. Lett.*, Vol. 19, 169-172.
- Fisher, M. E. & Barber, M. N. (1972). Scaling theory for finite-size effects in the critical region. *Phys. Rev. Lett.*, Vol. 28, 1516-1519.
- Fisher, D. S. (1986). Scaling and critical slowing down in random-field Ising systems. *Phys. Rev. Lett.*, Vol. 56, 416-419.
- For a review, see Folk, R.; Holovatch, Yu. & Yavors'kii, T. (2003). Critical exponents of a three dimensional weakly diluted quenched Ising model. *Usp. Fiz. Nauk*, Vol. 173, 175-200 [Phys. Usp., Vol. 46, 169-191].
- Folk, R. & Moser, G. (2006). Critical dynamics: a field-theoretical approach. *J. Phys. A*, Vol. 39, R207-R313.
- Gasparini, F. M.; Kimball, M. O.; Mooney, K. P.; & Diaz-Avila, M. (2008). Finite-size scaling of He4 at the superfluid transition. *Rev. Mod. Phys.*, Vol. 80, 1009-1059.

- Glauber, R. J. (1963). Time-dependent statistics of the Ising model. *J. Math. Phys.*, Vol. 4, 294-307.
- Gong, S.; Zhong, F.; Huang, X. & Fan, S. (2010). Finite-time scaling via linear driving. *New J. Phys.*, Vol. 12, 043036.
- Grassberger, P. (1995). Damage spreading and critical exponents for "model A" Ising dynamics. *Physica A*, Vol. 214, 547-559.
- Grinstein, G. (1976). Ferromagnetic phase transitions in random fields: The breakdown of scaling laws. *Phys. Rev. Lett.*, Vol. 37, 944-947.
- Halperin, B. I. & Hohenberg, P. C. (1967). Generalization of scaling laws to dynamical properties of a system near its critical point. *Phys. Rev. Lett.*, Vol. 19, 700-703.
- Harris, A. B. (1974). Effect of random defects on the critical behaviour of Ising models. *J. Phys. C*, Vol. 7, 1671-1692.
- Hasenbusch, M.; Pelissetto, A. & Vicari, E. (2007). Relaxational dynamics in 3D randomly diluted Ising models. *J. Stat. Mech.*, Vol. 2007, P11009.
- Hohenberg, P. C. & Halperin, B. I. (1977). Theory of dynamic critical phenomena. *Rev. Mod. Phys.*, Vol. 49, 435-479.
- Huang, X.; Gong, S.; Zhong, F. & Fan, S. (2010). Finite-time scaling via linear driving: Application to the two-dimensional Potts model. *Phys. Rev. E*, Vol. 81, 041139.
- Hui K. & Berker, A. N. (1989). Random-field mechanism in random-bond multicritical systems. *Phys. Rev. Lett.*, Vol. 62, 2507-2510.
- Hukushima, K. & Nemoto, K. (1995). On the forced oscillator method for the eigenvalue spectrum edge of  $\pm J$  random matrix. *J. Phys. Soc. Japan*, Vol. 64, 1863-1865.
- Ising, E. (1925), Beitrag zur Theorie des Ferromagnetismus. *Z. Phys.*, Vol. 31, 253-258.
- Jacobsen, J. L. (2000). Multiscaling of energy correlations in the random-bond Potts model. *Phys. Rev. E*, Vol. 61, R6060-R6063.
- Jacobsen, J. L. & Picco, M. (2000). Large- $q$  asymptotics of the random-bond Potts model. *Phys. Rev. E*, Vol. 61, R13-R16.
- Janssen, H. K.; Schaub, B. & Schmittmann, B. (1989). New universal short-time scaling behaviour of critical relaxation processes. *Z. Phys. B*, Vol. 73, 539-549.
- Janssen, H. K. (1992). On the renormalized field theory of nonlinear critical relaxation. *From Phase Transitions to Chaos*. Györgyi, G.; Kondor, I.; Sasvári, L. & Tél, T. (Ed.), World Scientific, Singapore.
- Kikuchi, M. & Ito, N. (1993). Statistical dependence time and its application to dynamical critical exponent. *J. Phys. Soc. Japan*, Vol. 62, 3052-3061.
- Kinzel, W. & Domany, E. (1981). Critical properties of random Potts models. *Phys. Rev. B*, Vol. 23, 3421-3434.
- Landau, D. P & Binder, K. (2005). *A Guide to Monte Carlo Simulations in Statistical Physics*, 2nd edition, Cambridge University Press, Cambridge.
- Li, Z. B.; Schülke, L. & Zheng, B. (1995). Dynamics Monte Carlo measurement of critical exponents. *Phys. Rev. Lett.*, Vol. 74, 3396-3399.
- Ma, S. -k. (1976). *Modern Theory of Critical Phenomena*, Benjamin, New York.
- Martin, P. C.; Siggia, E. D. & Rose, H. A. (1973). Statistical dynamics of classical systems. *Phys. Rev. A*, Vol. 8, 423-437.
- Mercaldo, M. T.; Anglès d'Auriac, J.-Ch. & Iglói, F. (2005). Disorder-driven phase transitions of the large  $q$ -state Potts model in three dimensions. *Europhys. Lett.*, Vol. 70, 733-739.

- Mercaldo, M. T.; Anglès d'Auriac, J.-Ch. & Iglói, F. (2006). Critical and tricritical singularities of the three-dimensional random-bond Potts model for large  $q$ . *Phys. Rev. E*, Vol. 73, 026126.
- Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller A. M. & Teller, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.*, Vol. 21, 1087-1092.
- Müller-Krumbhaar, H. & Binder, K. (1973). Dynamic properties of the Monte Carlo method in statistical mechanics. *J. Stat. Phys.*, Vol. 8, 1-24.
- Murtazaev, A.; Kamilov, I. & Babaev, A. (2007). Investigation of the critical properties of the three-dimensional weakly diluted potts model. *Bulletin of the Russian Academy of Sciences: Physics*, Vol. 71, 1586-1588.
- Murtazaev, A.; Babaev, A. & Aznaurova, G. (2008). Investigation of the influence of quenched nonmagnetic impurities on phase transitions in the three-dimensional Potts model. *Physics of the Solid State*, Vol. 50, 733-739.
- Newman, M. E. J. & Barkema, G. T. (1999). *Monte Carlo Methods in Statistical Physics*, Clarendon, Oxford.
- Nightingale, M. P. & Blöte, H. W. J. (1996). Dynamic exponent of the two-dimensional Ising model and Monte Carlo computation of the subdominant eigenvalue of the stochasticmatrix. *Phys. Rev. Lett.*, Vol. 76, 4548-4551.
- Nightingale, M. P. & Blöte, H. W. J. (2000). Monte Carlo computation of correlation times of independent relaxation modes at criticality. *Phys. Rev. B*, Vol. 62, 1089-1101.
- Okano, K.; Schulke, L.; Yamagishi, K. & Zheng, B. (1997). Universality and scaling in short-time critical dynamics. *Nucl. Phys. B*, Vol. 485 [FS], 727-746.
- Parisi, G.; Ricci-Tersenghi, F. & Ruiz-Lorenzo, J. J. (1999). Universality in the off-equilibrium critical dynamics of the three-dimensional diluted Ising model. *Phys. Rev. E*, Vol. 60, 5198-5201.
- Pazmandi, F.; Scalettar, R. T. & Zimanyi, G. T. (1997). Revisiting the theory of finite size scaling in disordered systems:  $\nu$  can be less than  $2/d$ . *Phys. Rev. Lett.*, Vol. 79, 5130-5133.
- Pelissetto, A. & Vicari, E. (2000). Randomly dilute spin models: A six-loop field-theoretic study. *Phys. Rev. B*, Vol. 62, 6393-6409.
- Pelissetto, A. & Vicari, E. (2002). Critical phenomena and renormalization-group theory. *Phys. Rep.*, Vol. 368, 549-727.
- Picco, M. (1998). A study of cross-over effects for the 2D random bond Potts model. e-print cond-mat/9802092.
- Potts, R. B. (1952). Some generalized order-disorder transformations. *Proc. Cambridge. Phil. Soc.*, Vol. 48, 106-109.
- (1990) *Finite Size Scaling and Numerical Simulations of Statistical Systems*, Privman, V. (Ed.), World Scientific, Singapore.
- Schehr G. & Paul, R. (2005). Universal aging properties at a disordered critical point. *Phys. Rev. E*, Vol. 72, 016105.
- Shima, H. & Nakayama, T. (1998). Finite-time scaling approach for the ac conductivity near the Anderson Transition. *J. Phys. Soc. Japan*, Vol. 67, 2189-2192.
- Shima, H. & Nakayama, T. (1999). Critical behavior of ac conductivity near the Anderson transition. *Phys. Rev. B*, Vol. 60, 14066-14071.
- Shinomoto, S. & Kabashima, Y. (1991). Finite time scaling of energy in simulated annealing. *J. Phys. A*, Vol. 24, L141-L144.
- Stanley, H. E. (1971). *Introduction to Phase Transitions and Critical Phenomena*, Oxford, London.
- Suzuki, M. (1974). New universality of critical exponents. *Prog. Theor. Phys.*, Vol. 51, 1992-1993.

- Suzuki, M. (1977). Static and dynamic finite-size scaling theory based on the renormalization group approach. *Prog. Theor. Phys.*, Vol. 58, 1142-1150.
- 't Hooft, G. & Veltman, M. (1972). Regularization and renormalization of gauge fields. *Nucl. Phys. B*, Vol. 44, 189-213.
- Tang S. & Landau, D. P. (1987). Monte Carlo study of dynamic universality in two-dimensional Potts models. *Phys. Rev. B*, Vol. 36, 567-573.
- Wansleben, S. & Landau, D. P. (1991). Monte Carlo investigation of critical dynamics in the three-dimensional Ising model. *Phys. Rev. B*, Vol. 43, 6006-6014.
- Wegner, F. W. (1972). Corrections to scaling laws. *Phys. Rev. B*, Vol. 5, 4529-4536.
- Weinberg, S. (1973). New approach to the renormalization group. *Phys. Rev. D*, Vol. 8, 3497-3509.
- Wiseman S. & Domany, E. (1995). Lack of self-averaging in critical disordered systems. *Phys. Rev. B*, Vol. 52, 3469-3484.
- Wiseman S. & Domany, E. (1998). Self-averaging, distribution of pseudocritical temperatures, and finite size scaling in critical disordered systems. *Phys. Rev. B*, Vol. 58, 2938-2951.
- Wu, F. Y. (1982). The Potts model. *Rev. Mod. Phys.*, Vol. 54, 235-268.
- Xiong, W.; Zhong, F.; Yuan, W. & Fan, S. (2010). Critical behavior of three-dimensional random-bond Ising model using finite-time scaling with extensive Monte Carlo renormalization-group method. *Phys. Rev. E*, Vol. 81, 051132.
- Xiong, W.; Zhong, F. & Fan, S. (2010). Positive specific-heat critical exponent of the three-dimensional random-bond Potts model. submitted.
- Yin, J. Q.; Zheng, B. & Trimper, S. (2004). Critical behavior of the two-dimensional random-bond Potts model: A short-time dynamic approach. *Phys. Rev. E*, Vol. 70, 056134.
- Yin, J. Q.; Zheng, B. & Trimper, S. (2005). Dynamic Monte Carlo simulations of the three-dimensional random-bond Potts model. *Phys. Rev. E*, Vol. 72, 036122.
- Yin, J. Q.; Zheng, B.; Prudnikov, V. V. & Trimper, S. (2006). Short-time dynamics and critical behavior of three-dimensional bond-diluted Potts model. *Eur. Phys. J. B*, Vol. 49, 195-203.
- Zhang, J. X.; Fung, P. C. W. & Zeng, W. G. (1995). Dissipation function of the first-order phase transformation in solids via internal-friction measurements. *Phys. Rev. B*, Vol. 52, 268-277.
- Zhang, J. B.; Wang, L.; Gu, D. W.; Ying, H. P. & Ji, D. R. (1999). Monte Carlo study of critical scaling and universality in non-equilibrium short-time dynamics. *Phys. Lett. A*, Vol. 262, 226-233.
- Zhang, J. X.; Zhong, F. & Siu, G. G. (1996). The scanning-rate dependence of energy dissipation in first-order phase transition of solids. *Solid State Commun.* Vol. 97, 847-850.
- Zheng, B. (1998). Monte Carlo simulations of short-time critical dynamics. *Int. J. Mod. Phys. B*, Vol. 12, 1419-1484.
- Zhong, F. (2002). Monte Carlo renormalization group study of the dynamic scaling of hysteresis in the two-dimensional Ising model. *Phys. Rev. B*, Vol. 66, 060401(R)
- Zhong, F. (2006). Probing criticality with linearly varying external fields: Renormalization group theory of nonequilibrium critical dynamics under driving. *Phys. Rev. E*, Vol. 73, 047102
- Zhong, F. & Chen, Q. Z. (2005). Theory of the dynamics of first-order phase transitions: Unstable fixed points, exponents, and dynamical scaling. *Phys. Rev. Lett.*, Vol. 95, 175701.

- Zhong, F.; Zhang, J. X. & Siu, G. G. (1994). Dynamic scaling of hysteresis in a linearly driven system. *J. Phys.: Condens. Matter*, Vol. 6, 7785-7796.
- Zhong, F.; Zhang, J. X. & Liu, X. (1995). Scaling of hysteresis in the Ising model and cell-dynamical systems in a linearly varying external field. *Phys. Rev. E*, Vol. 52, 1399-1402.
- Zhong, F.; Dong, J. M. & Xing, D. Y. (1998). Scaling of hysteresis in pure and disordered Ising models: Comparison with experiments. *Phys. Rev. Lett.*, Vol. 80, 1118-1118.
- Zhong, F. & Zhang, J. X. (1995). Scaling of thermal hysteresis with temperature scanning rate. *Phys. Rev. E*, Vol. 51, 2898-2901.
- Zhong, F. & Zhang, J. X. (1995). Renormalization group theory of hysteresis. *Phys. Rev. Lett.*, Vol. 75, 2027-2030.
- Zhong, F. & Zhang, J. X. (1997). Dynamic scaling of hysteresis in the 2D Ising model with impurities. *Acta Phys. Sin.*, Vol. 46, 791-795.
- Zhong, F. & Xu, Z. F. (2005). Dynamic Monte Carlo renormalization group determination of critical exponents with linearly changing temperature. *Phys. Rev. B*, Vol. 71, 132402.
- Zinn-Justin, J. (1996). *Quantum Field Theory and Critical Phenomena*, 3rd edn, Oxford, Clarendon.

# Using Monte Carlo Method to Study Magnetic Properties of Frozen Ferrofluid

Tran Nguyen Lan and Tran Hoang Hai  
*HoChiMinh City Institute of Physics,  
Vietnamese Academic of Science and Technology  
Viet Nam*

## 1. Introduction

Magnetic nanoparticles are single-domain particles of ferromagnetic or ferrite materials. Recently, magnetic nanoparticles have been applied more and more in technology, such as spintronics, magnetic recording, catalyst, and biomedicine. Therefore, experimental and theoretical studies on their magnetic properties are very important to provide essential information for individual applications. In addition, technical preparations have been developed fast, such as chemical synthesis, sputtering, or lithography. Depending on characters of assemblies, magnetic properties are different, such as two- or three-dimension, metallic or metallic oxide materials, order or disorder arrangement, surrounded by solids (granular solids) or liquids (ferrofluids), and the magnetic or non-magnetic surrounding matrix. This leads to studying fundamental properties of magnetic nanoparticle assemblies becomes interesting.

Among applicable potentials of magnetic nanoparticles, the biomedicine is a promising area, because the magnetic nanoparticles offer some great possibilities (Pankhurst et al., 2003). First, their size ranges from a few nanometers up to tens of nanometers. This means that their size can be smaller than or comparable to the size of biological entities, for example, a cell (10 – 100  $\mu\text{m}$ ), a virus (20 – 450 nm), a protein (5 – 50 nm) or a gene (2 nm wide and 10 – 100 nm long). Thus, they can penetrate easily into these entities. Second, these particles behave magnetic properties, so they can be controlled by an external magnetic field gradient. This opens application including the transport (drug delivery, cell separation) or immobilization (hyperthermia, contrast agent). Third, these particles can strongly resonate to a radio field. This makes them be easily excited by radio field leading, for example, heating in hyperthermia or magnetic resonance in contrast agent.

A key quality to study magnetic properties of particle is magnetic anisotropy energy (MAE). However, it is very difficult to exactly observe the MAE of each particle in the assembly. We can just obtain the MAE distribution of the assembly. Usually, the MAE distribution  $f(E_B)$  is deduced from the size distribution  $f(V)$  due to simplest expression of MAE,  $E_B = KV$ , with  $K$  and  $V$  as anisotropy constant and volume, respectively, of each particle. However, this way does not describe the exact information of real systems. It is due to some reasons as follow.

- (i) The size distribution obtains from microscopy images may not coincide with the real sample.
- (ii) The magnetic anisotropy involves many complexities, such as surface, magneto-

crystal, or shape. (iii) The orientation of anisotropy axis of particles in the assemblies is random, thus this way can not give the precise response between the size and the energy barrier distribution. A recent review written by Zheng et al. (Zheng et al., 2009) showed that there are some different ways to extract the anisotropy distribution, however, for the dilute sample. The problem becomes much more complex as the inter-particle interactions arise, namely dipolar interaction (for example, Bottoni et al., 1993; Ceylan et al., 2005; Parker et al., 2008) exchange interaction due to the contact between surface of particles or the magnetic surrounding matrix (for recent example, Tamion et al., 2010; Malik et al., 2010). However, the numerical analyses or phenomenal theory has not provided sufficient explanations for the observations of experiments. Therefore, computer simulations, especially Monte Carlo simulation, become efficient.

Now, to clearly see the successes of the Monte Carlo (MC) simulation of magnetic nanoparticle assemblies, we will shortly list some the important results. Kechrakos & Trohidou (Kechrakos & Trohidou, 1998) employed the MC simulation to give a general view about the interacting assemblies. Following these results, interacting assemblies possesses the anti-ferromagnetic state (decrease of magnetic responses) and ferromagnetic state (increase of magnetic responses) at low and high temperature, respectively. At the same time, MC method was used to seek the spin-glass (SG) like behavior of interacting systems at the low temperature. While almost results show the SG like behavior (for example, Anderson et al., 1997; Ulrich et al., 2003; Iglesias and Labarta, 2004; Fernandez and Alosa, 2009; ...), a few other results opposed the presence of SG like behavior (Garcia-Otero et al., 2000; Porto, 2005). Recent results based on the combination between the magnetic force microscopy and the MC simulation proved the existence of the short-range magnetic order deduced by dipolar interaction at the high temperature (Georgescu et al., 2006, 2008). These results visually asserted the role enhancing the energy barrier of anisotropic character along the bond axes of dipolar interaction. In addition, MC simulation is also a cost tool to investigate the structure of particle-cluster in assemblies (for a most recent example, Prokopieva et al., 2009). In this chapter, we will review some our recent results on magnetic properties of frozen ferrofluids.

Our chapter is organized as follow. In the part 2, we present two models describing the magnetic properties of dilute sample and the interacting sample. Therefore, we can see the limitation of these phenomenal models. In the third part, there are three section are presented, the field dependence of the blocking temperature, the concentration dependence of the coercive field, and the effective anisotropy distribution deduced by dipolar interaction. These results are clearly explained and compared to found experiments. Finally, we provide a short conclusion as well as some future aspects.

## 2. Some phenomenal models

### 2.1 Neel - Brown model

We first discuss the Neel - Brown (NB) model (Neel, 1953; Brown, 1963) which is used to describe the magnetic properties of assemblies in the non-interacting case. In this model, the relaxation of each particle moment in the assemblies depends on the competition between the thermal fluctuation  $k_B T$  and the barrier energy  $E_B$ , which defined by the magnetic anisotropy of each particle. First assumption of this model is that each particle moment is formed by the rigid alignment of atomic spins; therefore, the reversal of particle moments is

the coherent rotation of atomic spins. The relaxation time, without external field, is characterized by the Arrhenius law as follow

$$\tau = \tau_0 \exp\left(\frac{E_B}{k_B T}\right) \quad (1)$$

Where  $\tau$  is relaxation time and  $\tau_0$  is characteristic time which is a function of gyro-magnetic ratio, longitudinal magnetostriction constant, and Young modulus. Neel estimated  $\tau_0$  to be of order  $10^{-10}$  s, this value is good agreement with experiments. Following the expression (1), if  $k_B T \ll E_B$  the  $\tau$  is so large (very slow relaxation) that relaxation can not be observed, that is the assembly seems to be an ordered magnetic system. On the contrary, if  $k_B T \gg E_B$  the relaxation is very fast, so the assembly reach to the thermal dynamic equilibrium. Therefore, there is a finite temperature which separates two above regimes. This temperature is called the blocking temperature and determined from the expression (1)

$$T_B = \frac{E_B}{k_B \ln(\tau_m / \tau_0)} \quad (2)$$

with  $\tau_m$  as the measured time-window. Clearly that  $T_B$  strongly depends on the  $\tau_m$  which is defined by experimental conditions, so the blocking temperature is not unique.

For  $T < T_B$ , magnetic moments can not reverse and the assembly is in the blocking state exhibiting hysteresis. For  $T > T_B$ , the magnetic moments easily reverse, the hysteresis disappears and the assembly is in the super-paramagnetic states (SPM). In the case of uniaxial anisotropy barrier with anisotropy constant as  $K_u$ ,  $E_B = K_u V$ , the blocking temperature has form

$$T_B = \frac{K_u V}{k_B \ln(\tau_m / \tau_0)} \quad (3)$$

In the poly-dispersity sample, the blocking temperature of sample is an average of all blocking temperatures of individual particles. Namely

$$\langle T_B \rangle = \frac{\int T_B(V) f(V) dV}{\int f(V) dV} \quad (4)$$

with  $\langle T_B \rangle$  as the blocking temperature of sample.

In the presence of the external field, the anisotropy barrier is reduced,  $E_B = K_u V (1 - H/H_a)^\alpha$ , with  $H_a = 2K_u/M_s$  as the anisotropy field,  $M_s$  as saturated magnetization and the parameter  $\alpha$  closes to 1.5 (Knobel et al., 2008) or 2 (Kechrakos, 2010). The blocking temperature of each particle has form

$$T_B = \frac{K_u V}{k_B \ln(\tau_m / \tau_0)} \left[ 1 - \frac{H}{H_a} \right]^\alpha \quad (5)$$

with  $T_B(0)$  as the blocking temperature of each particle in the absence of external field. Following the expression (5), the blocking temperature would monotonic decrease with

increasing the external field. However, the experimental results showed a non-monotonic field dependence of blocking temperature in the dilute sample. We will explain this contrary by Monte Carlo simulation in the fourth part. Effect of inter-particles interactions on this dependence will be also provided.

## 2.2 RAM model for interacting particle assemblies

Now, we will briefly consider a recent approach which was developed by Nunes et al. (Nunes et al. 2005). This approach was based on the random anisotropy model (RAM) which was developed to describe the magnetic properties of amorphous magnetic materials. According to RAM model, the correlation length due to interactions between particles is included in the effective value of anisotropy constant  $K_{eff}$ . The aim of this approach is to calculate the blocking temperature under the influence of the interactions in the presence of external field. Therefore, the authors began from the expression (5) with the two relevant parameters which need to be modified, namely effective anisotropy and effective volume  $V_{eff}$  of particles in the correlation length  $L$  (Nunes et al., 2005)

$$K_{eff} = \frac{K}{\sqrt{N}} \quad \text{and} \quad V_{eff} = \frac{\pi}{6} \left[ D^3 - x(L^3 - D^3) \right] \quad (6)$$

Where  $N$  is number of correlated particles and  $D$  is the diameter of each particle,

$$N = \left[ 1 + x \frac{(L^3 - D^3)}{D^3} \right] \quad (7)$$

As the interaction is weak,  $L \ll D$ , both expressions in (6) will tend to the anisotropy and volume of individual particles.

Next, the authors gave the correlation length as a function of external field

$$L_H = D + \sqrt{\frac{2A_{eff}}{M_s H_{dc} + C}} \quad (8)$$

with  $A_{eff}$  represents the intensity of interactions and  $C$  is a parameter to prevent the divergence at zero fields.

By substituting the effective anisotropy and effective volume of each particle in expression (6) on the expression (5), the authors obtained the expression for the blocking temperature of couple particles in the term of structural parameter (Nunes et al., 2005) as follow

$$T_B = \frac{K_u \pi \left[ D^3 - x(L_H^3 - D^3) \right]}{6k_B \ln(\tau_m / \tau_0) \left[ 1 + x \frac{(L_H^3 - D^3)}{D^3} \right]^{1/2}} \left\{ 1 - \frac{H_{dc}}{H_a} \left[ 1 + x \frac{(L_H^3 - D^3)}{D^3} \right]^{1/2} \right\}^a \quad (9)$$

This phenomenal approach obtained several good agreements with experiments of granular solids (Nunes et al., 2005; Knobel et al., 2008). However, it has not still given a general view on collective states of strongly interacting magnetic nanoparticle systems (Knoble et al., 2008).

### 3. Energy and simulation method

#### 3.1 Energy

In the computer simulation, energy function is an extremely important problem. Therefore, the first work in the computer simulation is to build an energy function which must relate to real systems. Our model is based the earlier studies on Monte Carlo simulation of magnetic nanoparticle systems (for example, Kechrakos and Trohidou, 1998; Garcia-Otero et al., 2000), thus some assumptions are used to introduce the energy as follow

- i. Particles possess the totally spherical shape with diameter  $D$ . The poly-dispersity of particle size is determined by the log-normal distribution with the width  $\sigma < 1$ .
- ii. The magnetic moment vector of each particle has form  $\boldsymbol{\mu} = M_s V \mathbf{e}$ . Where  $M_s$ ,  $V$ , and  $\mathbf{e}$  are saturated magnetization, volume, and the unit vector of each magnetic moment, respectively.
- iii. Position of particles is randomly arranged in the cubic box of the volume  $(n.D)^3$  with  $n$  as an integer.
- iv. The anisotropy of each particle is characterized by uniaxial anisotropy with the anisotropy constant  $K_u$  being in the range from  $19 \text{ kJ.m}^{-3}$  to  $190 \text{ kJ.m}^{-3}$ .
- v. Thermal fluctuations of the assembly magnetization are well described by the coherent rotation of magnetic moments of particles.

Basing on the above assumptions, we obtain the total energy function for each particle

$$E^{(i)} = -K_u V_i \left( \frac{\boldsymbol{\mu}_i \cdot \mathbf{n}_i}{|\boldsymbol{\mu}_i|} \right)^2 - \boldsymbol{\mu}_i \cdot \mathbf{H} + g \sum_{j \neq i}^N \left( \frac{\boldsymbol{\mu}_i \cdot \boldsymbol{\mu}_j}{r_{ij}^3} - 3 \frac{(\boldsymbol{\mu}_i \cdot \mathbf{r}_{ij})(\boldsymbol{\mu}_j \cdot \mathbf{r}_{ij})}{r_{ij}^5} \right) \quad (10)$$

The first term in Eq.10 is the anisotropy energy,  $\mathbf{n}_i$  is the direction of the anisotropy axis,  $|\mathbf{n}_i| = 1$ . The second term is the Zeeman energy,  $\mathbf{H}$  is the external field. The last time is the dipolar energy between two particles  $i$  and  $j$  separated by  $r_{ij}$ , and constant  $g = \mu_0/4\pi$ .

Usually, in the mean field theory (Strikmann and Wohlfarth, 1981), the effect of dipolar interaction is represented through the dipolar field which is included into the applied field. However, following the DBS model (Dormann et al., 1999) and some simulated as well as experimental results (for example, Kechrakos and Trohiou, 1998; Verdes et al., 2002; Ceylan et al., 2005; Knobel et al., 2008; ...) showed that the dipolar interaction deduces the effect being similar to anisotropy barrier, namely the enhancement of the blocking temperature along with the interacting strength. Therefore, we can include the anisotropic character along bond axes of the dipolar interaction in the anisotropy energy.

1. Case of the inclusion of dipolar field on the applied field, the dipolar field is defined by follow equation

$$\mathbf{H}_{dipol}^i = - \sum_{j \neq i}^N \frac{\partial U_{dipol}^{ij}}{\partial \boldsymbol{\mu}_i} = -g \sum_{j \neq i}^N \left( \frac{\boldsymbol{\mu}_j}{r_{ij}^3} - 3 \frac{\mathbf{r}_{ij} \cdot (\boldsymbol{\mu}_j \cdot \mathbf{r}_{ij})}{r_{ij}^5} \right) \quad (11)$$

Then, the dipolar energy of the particle  $i$  can rewrite in the simple form  $U_{dipol}^i = \boldsymbol{\mu}_i \cdot \mathbf{H}_{dipol}^i$ . And the energy of the particle  $i$  as

$$E^{(i)} = -K_u V_i \left( \frac{\boldsymbol{\mu}_i \cdot \mathbf{n}_i}{|\boldsymbol{\mu}_i|} \right)^2 - \boldsymbol{\mu}_i \cdot \mathbf{H}_i^{eff} \quad (12)$$

Now, the system can be thought as an ensemble of the non-interacting particles feeling an effective field that is sum of an external and a local field  $\mathbf{H}_i^{eff} = \mathbf{H} + \mathbf{H}_{dipol}^i$ .

2. Case of the inclusion of anisotropic character along bond axes of the dipolar interaction on the uniaxial anisotropy, we rewrite the total energy of each particle

$$E^{(i)} = -K_{eff}^i V_i + gM_S^2 \sum_{j \neq i}^N \frac{\mathbf{e}_i \cdot \mathbf{e}_j}{r_{ij}^3} V_j - M_S V_i \mathbf{e}_i \cdot \mathbf{H} \quad (13)$$

The first term in Eq.13 is the effective anisotropy energy with effective anisotropy density

$$K_{eff}^i = K_u (\mathbf{e}_i \cdot \mathbf{n}_i)^2 + 3gM_S^2 \sum_{j \neq i}^N \frac{(\mathbf{e}_i \cdot \mathbf{r}_{ij})(\mathbf{e}_j \cdot \mathbf{r}_{ij})}{r_{ij}^5} V_j \quad (14)$$

The second term in Eq. 13 characterizes the anti-ferromagnetism because of  $gM_S^2 > 0$  and it just has significance for N neighbour particles. While the second term in Eq. 14 expresses the ferromagnetic anisotropy along the bond axis. And the effective anisotropy energy of each particle has form

$$E_{Beff}^{(i)} = K_{eff}^i V_i = \left[ K_u (\mathbf{e}_i \cdot \mathbf{n}_i)^2 + 3gM_S^2 \sum_{j \neq i}^N \frac{(\mathbf{e}_i \cdot \mathbf{r}_{ij})(\mathbf{e}_j \cdot \mathbf{r}_{ij})}{r_{ij}^5} V_j \right] V_i \quad (15)$$

In our opinion, two above inclusions have different advantages as considering the effect of the dipolar interaction at the different temperature ranges. If the temperature is low, the thermal fluctuation of magnetic moment is very weak. This means that the effect of the dipolar field deduced by a magnetic moment on other moments is very significant. On the other hand, if the temperature is high to make the magnetic moment fluctuate, the dipolar field continuously varied. Therefore, the anisotropic character which enhances the anisotropy barrier becomes more important as we will explain later. In summary, the first inclusion should be used to explain the collective state at the low-temperature and the second one more clearly shows the role enhancing the anisotropy barrier. These inclusions are similar to the mean field theories (Walton, 2007; Dotsenko, 2010) and DBF model (Dormann et al., 1988); however, with computer simulation the interaction field and the effective anisotropy energy are exactly calculated, because their values totally obtain through the configurations of all the moments in the system and they change in each time-step (Kechrakos, 2010). Finally, note that although the applied field was not showed in the Eq.15, its influence is implied through the direction of each magnetic moment.

### 3.2 Simulation procedure

Monte Carlo method is an easy and fast approach to minimize the energy of random or pseudo-random systems. With the magnetic nanoparticle systems the Metropolis algorithm (Metropolis et al., 1953) is most widely used. Details of this algorithm have been presented in many previous literatures, so in here we just introduce the procedure to obtain the local energy minima of each particle in the assembly. The configurations of magnetic moments are performed in the spherical coordinate. We assume that the external field  $\mathbf{H}$  is applied along the z-axis of the particle system, and easy axis of particles aligned at an angle  $\psi$  with the field and the direction of the magnetic moment is determined by values  $(\theta, \varphi)$ . At the

beginning of each simulation, an assemblage of  $N$  particles is generated and random values of  $\varphi$ , and  $\theta$  are drawn from a uniform distribution  $\theta \in [0, 2\pi]$ , and  $\varphi \in [0, 2\pi]$ . The  $\psi$  value of each particle is constant throughout the simulations and the variation of  $\varphi$  and  $\theta$  is of interest. Each Monte Carlo step consists of the following steps

- i. Using Eq. (10) the energy of each particle is determined base on the applied field and the current values of  $\theta, \varphi$ , this value is  $E$ .
- ii. A new orientation of the magnetization is selected at random within even angles ( $d\varphi$  and  $d\theta$ , these values are determined randomly from  $[-\eta_{\max}, \eta_{\max}]$ ).
- iii. The energy  $E_{\text{trial}}$  is calculated for the particle along with the new values of the magnetization.
- iv. The difference  $\Delta E$  is calculated for the two possible orientations of the magnetization,  $\Delta E = E_{\text{trial}} - E$ .
- v. The magnetization of the particle is moved to the new orientation with the probability  $\min [1, \exp (-\Delta E/K_B T)]$ .

All simulations are performed by the dimensionless parameter of magnetization and applied field viz.  $m = M/M_S$  and  $h = H/H_a$ , respectively. Following our previous reports (Lan and Hai, 2010), the mean diameter of particles is 7.5 nm.

## 4. Our results and discussion

### 4.1 Field dependence of blocking temperature

According to the Neel – Brown model, the blocking temperature will monotonic decrease along with the applied field in the dilute samples. However, many experiments found that this dependence is non-monotonousness with various materials, such as  $\text{Fe}_3\text{O}_4$ , ferritin,  $\gamma\text{-Fe}_2\text{O}_3$ , Co and FePt (Luo et al., 1991; Friedmann et al., 1997; Sappey et al., 1997; Kachkachi et al., 2000; Zheng et al, 2006). On the contrary, the blocking temperature seems to be invariable at the low-field in the case of dense sample (Kachkachi et al., 2000; Parker et al., 2008). However, there have been no clear explanations by theoretical employment. Therefore, we use the Monte Carlo simulation to investigate these problems.

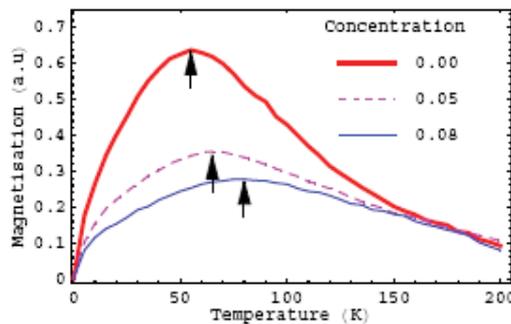


Fig. 1. Temperature dependence of magnetization as a function of concentration, the blocking temperatures (peaks of curves) are shown by arrow (from Lan and Hai 2010)

We define the blocking temperature of assembly as the peak temperature of the zero-field-cooling (ZFC) curve. Therefore, to find the blocking temperature we have to simulate the ZFC process. First the sample is cooled to very low temperature without external field. After that, a small field is applied and the sample is heated up to a very high temperature with the slow rate. The peak of the curve is the blocking temperature. At each the value of

temperature  $5.10^4$  Monte Carlo steps are used. The results are average of 100 samples with different initial configurations. The energy barrier energy is extracted to explain these results. It is very difficult to build a numerical model that finds the barrier distribution. However, we recognize that the energy difference  $\Delta E$  in the translation probability is always equal to one of the actual energy barriers of the system (Iglesias & Labarta, 2004). Therefore, we can extract the barrier distribution by using the Monte Carlo method to sample the individual energy barriers of all the particles. Fig. 1 represents the ZFC curve at the different concentration. The peak temperature shifts to the large-value along with the increase of concentration as found in very many previous studies.

Now, we consider the field dependence of the blocking temperature in two cases, dilute and dense sample

#### 4.1.1 Dilute sample

In the Fig. 2a, when the reduced field value is smaller than 0.3, the peak temperature increases along with the increase of the reduced field, and then it continuously decreases along with the increase of the reduced field. The increase of the peak temperature in the low field expresses strongly at the large- $\sigma$ . As saw in Fig. 2b, the distribution of the energy barrier is sensitive to the applied field, and they are broadened as the field increases at the low values. Sappey et al. (Sappey et al., 1997) suggested that the effect of low fields on the energy barrier distribution could be due to disorder of orientations, or the defects of each particle. Zheng et al. (Zheng et al., 2006) explained that this non-monotonous was due to combining the size distribution and the slow decrease of the magnetization (or non-Curie's law dependence of magnetization) above the blocking temperature in the field.

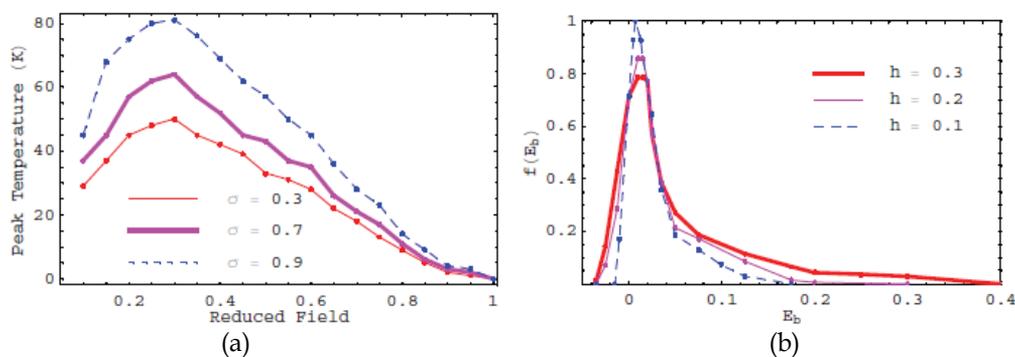


Fig. 2. a) Field dependence of blocking temperature (peak ZFC curve) and b) the barrier distribution as a function of the low field in the dilute sample (from Lan and Hai 2010)

Although these explanations satisfy with the present model (the uniaxial anisotropy model), they are simple and not sufficient. Recently, Perez et al. (Perez et al., 2008) showed the effect of the anisotropy ratio (core-shell anisotropy) on the energy barrier distribution of the dilute systems and with increasing the surface anisotropy the energy barrier distribution is enlarged even when the size distribution is quite narrow. Basing on this, we can imagine that under the influence of the low field, the contribution of the surface anisotropy may become dominant in the comparison with the core anisotropy, so the energy barrier distribution of the system is enlarged. At the high field, the Zeeman energy exceeds the anisotropy energy of each particle. Therefore, atomic moments of each particle containing the atomic surface and the atomic core orient along with the field direction, this means that

the disorder at the particle surface disappears. In other words, the contribution of the surface anisotropy is not dominant, so the barrier distribution becomes narrow, and the peak temperature decreases. In our opinion, the non-monotonic behavior of the  $T_p$  vs.  $h$  curvature expresses strongly in the independent particle systems possessing the strong surface anisotropy. We can use the atomistic simulation methods, such as the first principal calculation, to predict the influence of the applied field on the particle surface structure.

#### 4.1.2 Dense sample

We considered the influence of the dipolar interaction on the behavior of the ZFC-peak vs. applied field curve. A recent numerical study was based on the Gittlmen-Abeles-Bozowski model (Agzegagh & Kachkachi, 2007) to perform the change of the shape of this curve for the weak interaction, but it was complex and at the high concentrations, the numerical analyses are impossible. We need to remember that the dipolar energy of the particle  $i$  will make an effective anisotropy which makes the increase the blocking temperature. As we saw in Fig. 3a, the curvature change from the non-monotonousness to monotonousness, and at the very strong interaction, the curvature becomes flatter. As in Fig. 3b, when the interacting strength increases, the relation between the size distribution and the barrier distribution disappears, therefore, the non-monotonousness will change to the monotonousness. However, if the sample is very dense, the dipolar interaction is strong enough to remain the energy barrier at the small values of the applied field. Then the energy barrier of each particle slowly decreases along with the applied field, this means that the curvature is less sloping. Because the strong interaction remain the barrier as mentioned above, the blocking temperature exists even the applied field exceeds the anisotropy field  $H_a$ . This scenario also found by Serantes et al. (Serantes et al., 2008), however, the behaviors at the low field are rather different from our results. We find that the curvatures are separated clearly at the low field, because at this region the local field plays dominant role, then the blocking temperature increases along with the increase of the concentration.

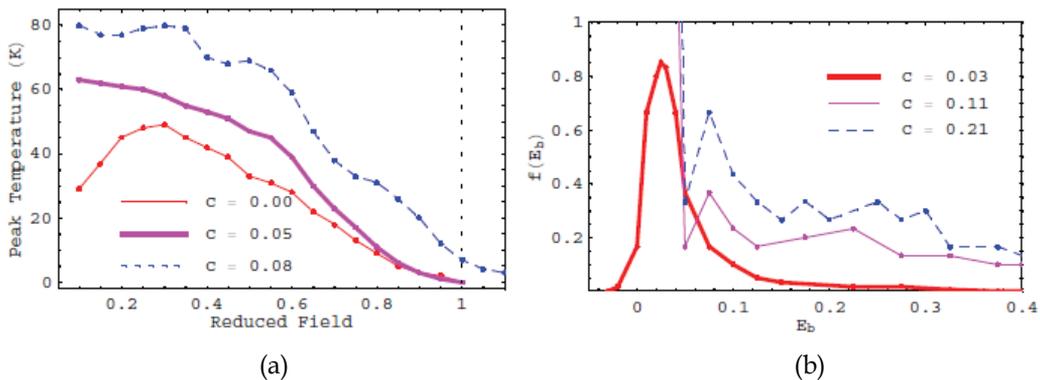


Fig. 3. a) Field dependence of blocking temperature (peak ZFC curve) and b) the barrier distribution as a function of the low field in the dense sample (from Lan and Hai 2010)

#### 4.2 Concentration dependence of coercive field

Coercive fields are obtained by using the procedure as follow. The sample is first submitted to equilibrium state through  $10^4$  Monte Carlo steps, and then an external field is applied and

increased until obtaining the saturation of magnetization. The configurations of magnetic moments are recorded and the external field is decreased to zero. The magnetizing process is performed again with the negative direction of the applied field. The coercive field is value at which the magnetization equals to zero. In each increasing (decreasing) step of field  $5 \cdot 10^4$  Monte Carlo steps are performed. The results are the average of 100 different samples of random initial configurations. Fig. 4 shows the hysteresis at the temperature  $T = 10$  K as a function of concentration.

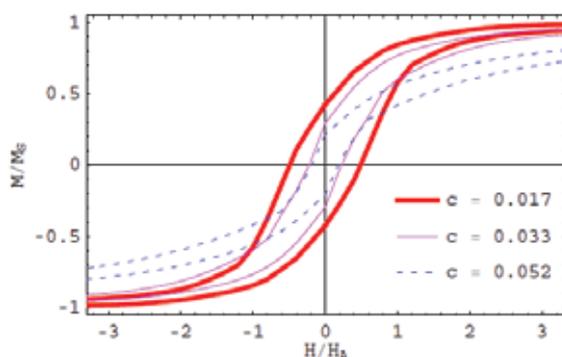


Fig. 4. Hysteresis loops at the low-field as a function of the concentration and the low-temperature is set at 10 K. The results are the average of 100 different samples of random initial configurations (alignment of magnetic moments and arrangement of positions) (from Lan and Hai 2010).

In fact, the concentration dependence of magnetic responses was studied in the previous simulated and experimental works (Bottoni et al., 1993; Kechrakos & Trohidou, 1998; Verdes et al., 2002; Blackwell et al., 2003; Ceylan et al., 2005; Tan et al., 2010). These results found a cusp of the magnetic response vs. concentration curves. However, there are differences in the explanations between the simulations and experiments. Experimental results (Bottoni et al, 1993; Ceylan et al, 2005) asserted that the increase of coercive field along with the concentration is due to the chain-like or small cluster formation which enhances the energy barrier under the influence of interparticle interactions. While the simulation results (Kechrakos and Trohidou, 1998; Verdes et al., 2002) showed that the origin of this cusp is the competition between the blocking and super-paramagnetic states below a concentrated threshold. Although these differences are due to in the real system the particle systematic formations of particle-clusters that do not take place in the simulations, we believe that the magnetic phases are also dominant in the appearance of this cusp. Therefore, to confirm the simulated results we do calculate on the poly-dispersity frozen ferrofluids. That is, the poly-dispersity will decide the competition of magnetic phases (blocking or super-paramagnetic state) of dilute sample at the finite temperature.

First, we consider the temperature dependence of coercivity to show the suitable finite temperature in the current model. Fig. 5a shows the temperature dependence of coercivity. As the temperature is very low, the coercive decreases along with the dipolar interacting strength and with rising the temperature, the coercive field slower decay at the higher concentration. These results also were performed by Kechrakos & Trohidou (Kechrakos & Trohidou, 1998) in the mono-dispersity system. As a result, at a finite temperature, the coercive field increases along with the concentration and we may call this temperature as

the transition temperature which separates the anti-ferromagnetic and ferromagnetic regime. It is worth that the transition temperature is not unique, it depends on the change of concentration. This unique behavior leads to a difficulty for determining the magnetic phase diagram of ferrofluid viz. at a same temperature, a sample can be in the anti-ferromagnetic states in comparison with any sample having the lower concentration but in the ferromagnetic regime with the another one having the higher concentration. It is interesting that in a sample, the transition temperature is always less than the peak temperature of the dc zero-field-cooling magnetization curve. This result was also experimentally found in the strong interacting particle system (For example, Kleemann et al, 2001; Parker et al., 2008).

Fig. 5b performs magnetic coercivity vs. concentration curves at the finite temperature,  $T = 50$  K (the dash line in Fig. 5a). The non-monotonousness expresses clearly in the mono-dispersity sample and the plateau-like shape appears with increasing the distribution width. If we continue raising the distribution width, the curves may monotonously decrease. As presented in the above result (sec. 4.1), with our system the blocking temperature of the mono-dispersity and weak interacting sample is about 50 K. Therefore, at the low concentration, the mono-dispersity sample has many super-paramagnetic particles, so as explained above the dipolar interaction deduces the increase of the magnetic response. On the contrary, in the strong poly-dispersity sample, the number of super-paramagnetic particles is small, therefore this weakens the non-monotonousness. With increasing the concentration, certainly, the dipolar interaction restores the demagnetizing role and then the coercivity decreases. Interestingly, magnetic responses at the low concentration completely separate, while they are close together at the high concentration. This situation again asserts the role exchange of the poly-dispersity and the dipolar interaction as discussed above. Clearly, the poly-dispersity also contributes to the complexity of the magnetic phase diagram of the magnetic nanoparticle system.

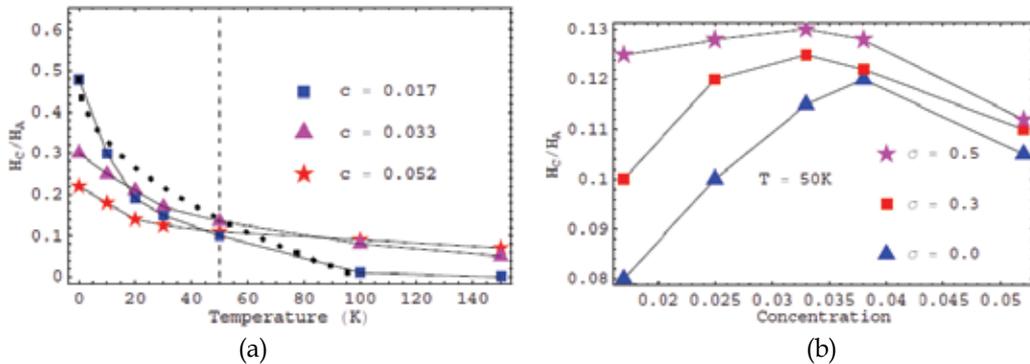


Fig. 5. (a) Temperature dependence of coercive field as a function of concentration and (b) concentration dependence of the coercive field as a function of distribution width  $\sigma$  at the finite temperature  $T = 50$  K (from Lan and Hai 2010)

#### 4.3 Effective anisotropy in interacting assemblies at the finite temperature

As saw above, at a finite temperature, the dipolar interaction causes the enhancement of energy barrier which makes the coercive field increase. A question in here is what deduces this behavior? We attribute above behaviors to the thermal fluctuation of particle moments.

In the blocking state, under the influence of the anti-ferromagnetic character of dipolar interaction (the second term in Eq. 13) magnetic moments will tend to anti-parallel behaviors to minimize the dipolar energy. However, with increasing the temperature, the thermal fluctuation makes the random orientation of particle moments, and the anti-ferromagnetic role of the second term in Eq. 13 seems to be vanished. While the position of the particle does not change, in the other word, the bond axes are conserved. This leads to the dominance of the anisotropic component along the bond axes in the dipolar interaction (the second term in the Eq. 14), therefore the effective anisotropy increases. These arguments are proved by moment snapshots in Fig. 6. Positions of particles are chosen randomly. Configurations are recorded after the temperature is slowly enhanced from zero to 100 K. As we see in Fig. 6, the particle moments tend to the alignment along the bond axes in the direction of external field  $\mathbf{H}$ .

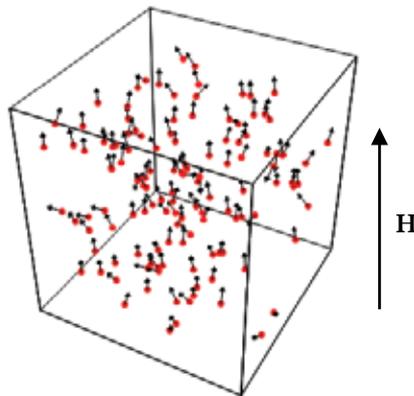


Fig. 6. Snapshot of 128 magnetic moment at the temperature  $T = 50$  K, small arrows represent magnetic moments and the large solid arrow indicates the direction of external field

At each the given condition, the sample is firstly submitted to equilibrium state through  $5.10^4$  Monte Carlo steps. Next, we calculate the energy barrier (EB) and the effective anisotropy (EA) of each particle. (i) EB is determined similarly to the energy difference in each Monte Carlo step, but the only positive value is accepted. (ii) EA is calculated from Eq. 15. As we mentioned, although the effect of the external field was not clearly included in Eq. 15, it is implicated in the direction of magnetic moments.

Recent experimental results (Georgescu et al., 2006) in the two-dimension particle array of  $\gamma$ - $\text{Fe}_2\text{O}_3$  particles showed that dipolar interaction deduces flux closure configurations along the bond axes which blocks magnetic moments. As we introduced, the effective anisotropy (EA) was determined by the uniaxial anisotropy and anisotropic character along the bond axes of dipolar interaction, so we do compare between the EB distribution and the EA distribution with the change of the particle number at the  $T = 100$  K as in Fig. 7. The uniaxial anisotropy constant is invariable,  $K_u = 19$  kJ.m<sup>-3</sup>. With the dilute sample, Fig. 7a,  $N = 64$ , the large-energy tails of the EA distribution and the EB distribution are strongly different, while the relative identicalness exists in the dense sample,  $N = 128$ , as in Fig. 7b. These lead to the assertion that the dipolar interaction dominates in the large-energy tails at the high-temperature as found in the experiment (Georgescu et al., 2006). In the later work, Georgescu et al. (Georgescu et al., 2008) showed that the role enhancing the energy barrier

of dipolar interaction strongly depends on temperature. This means that at the low temperature, the magnetic moments are blocked by influence of the intrinsic anisotropy. On the contrary, as the temperature is high enough the role of the intrinsic anisotropy become weak and the dipolar interaction dominate in the effective anisotropy which enhancing the blocking temperature. These results seem to respond to our argument in the section 4.2 about magnetic phase of interacting assemblies.

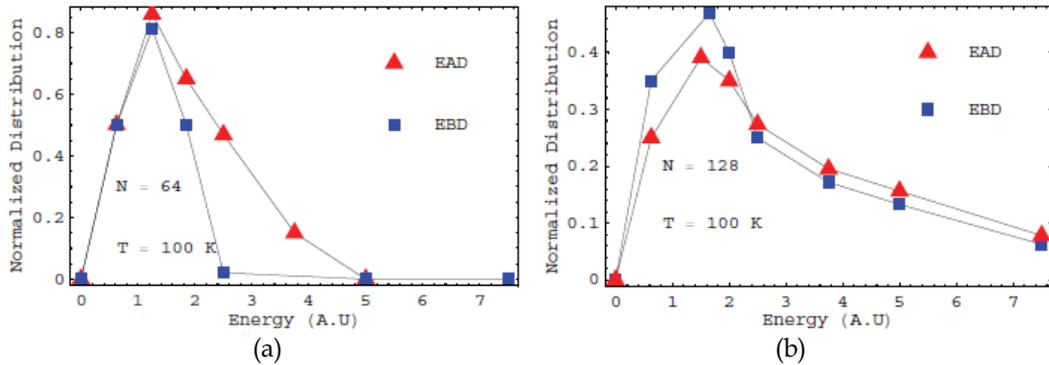


Fig. 7. Comparison between the effective anisotropy distribution (EAD) and the energy barrier distribution (EBD) at finite temperature  $T = 100$  K with the change of number of particles.

Finally, to show the effect of thermal fluctuation on the effective anisotropy deduced by the dipolar interaction, we perform the temperature dependence of the EA distribution as in Fig. 8. We can see that the distribution enlarges as the temperature increases. However, if the temperature is very high, the distribution becomes narrow. These can be explained as follow. At the low-temperature, magnetic moments have the anti-paralleling tendency to minimize the dipolar interacting energy. The effective anisotropy along the bond axis is therefore trivial. At the moderate temperature, the thermal fluctuation is small enough for weakness of the anti-ferromagnetic character of dipolar interaction and then the ferromagnetic character along the bond axis becomes important. However, at the very high temperature the strong thermal fluctuation makes the ferromagnetism disappearing.

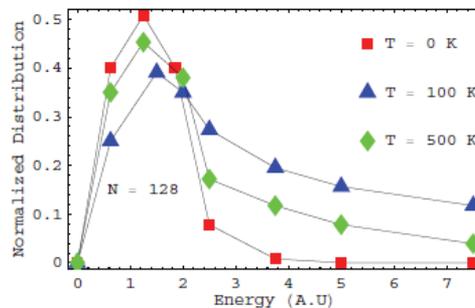


Fig. 8. Effective anisotropy distribution as a function of temperature

If temperature is continued increasing the sample behaves the paramagnetic character. These mean that EA distribution just has prime significance in a certain interval which is

around blocking temperatures. At very high temperature, the sample behaves the paramagnetic states. In the summary, the role of dipolar interaction strongly depends on the range of temperature.

## 5. Conclusions and future aspects

We presented several fundamental properties of frozen ferrofluids, which have not been sufficiently given by the phenomenal models, through the Monte Carlo simulation. These results are good conformity with found experiments.

- The field dependence of the blocking temperature behaves differences in the samples with the various concentrations. (i) The non-monotonousness in the dilute sample, which contrasts with the Neel - Brown theory, is due to the responses between the energy barrier distribution and the size distribution. These results showed effect of poly-dispersity on the blocking temperature in magnetic nanoparticle assemblies as found in the experiments. However, to clearly understand this non-monotonousness we may base on the recent results on the surface anisotropy of magnetic nanoparticles. (ii) The invariance of the blocking temperature along with the low-field in the dense sample is due to the role enhancing the anisotropy of the dipolar interaction. In other words, the dipolar interaction makes a new order state which behaves the ferromagnetism along the bond axes.
- The demagnetizing role of the dipolar interaction strongly depends on the temperature as well as the poly-dispersity. We had proven that the main cause of the increase of coercivity along with the low concentration is due to the competition between the number of blocked and super-paramagnetic particles through the poly-dispersity, therefore, we believe that only the dipolar interaction has significance in the non-monotonous variation of the coercive field vs. concentration curve. We can see that the magnetic phase diagram of frozen ferrofluid not only depends on the concentration but also is affected by the poly-dispersity. At a finite temperature, the strong poly-dispersity retains the anti-ferromagnetic role of the dipolar interaction and the contrary, this role weakly expose in mono-dispersity samples. Finally, depending on the type of materials, the cusp can occur even at the room temperature, so we can control the concentration or the poly-dispersity of the sample in the preparation to satisfy applicable potentials.
- We have clearly explained the cause of the collective state at the high-temperature in frozen ferrofluid through the effective anisotropy distribution: at a temperature being near the blocking temperature, the thermal fluctuation deduces the vanishing of effect of the magneto-crystal anisotropy, while the collective dipole-dipole interaction enhances the effective anisotropy by forming the flux closure configurations along bond axes. In addition, our results also showed that the anisotropic character along the bond axis of dipolar interaction plays dominant role in effective anisotropy at the interval of temperature being near blocking temperature. We can apply this relatively simple way to explain the collective states in the systems of different spatial arrangement

In our opinions, there are some suggestions on the theoretical model on magnetic properties and the simulations in biomedical applications as follow

- As we discussed above, there are two ways to include the dipolar interaction on the other terms, namely applied field and the uniaxial anisotropy. Therefore, depending on the range of temperature (low- or high-temperature), we select the suitable model. For

example, at the low temperature, we can develop mean field theory to describe the super-spin glass behaviors or at the high temperature, we can use the model of Dormann et al. 1988 to show the ferromagnetic states of the interacting assemblies, however with more exactly parameter. Clearly that no model can cover all complexities of interacting magnetic nanoparticle assemblies, so each model just describe properties of assemblies in the specific case.

- Recent simulated as well experimental studies on applications of collective particles in biomedicine (Schaller et al., 2009; Dennis et al., 2007, 2008) showed great promises. Therefore, we need to develop computer simulations to optimize as well as predict the applicable potentials of the collective state at the high temperature, such as effect of interactions on heating in hyperthermia, resonance of magnetic moment in contrast agents, the transport in drug delivery or cell separation. Besides, the collective particle assemblies may open new applications which we have to very much effort to find.

## 6. References

- Allia, P.; Coisson, M.; Tiberto, P.; Vinai, F.; Knobel, M.; Novak, M. A. & Nunes, W. C. Granular Cu-Co alloys as interacting superparamagnets. *Phys. Rev. B*, Vol. 64, (September, 2001), pp. 144420 (1-12), ISSN: 1098-0121
- Andersson, J.-O.; Djurberg, C.; Jonsson, T.; Svedlindh, P. & Nordblad, P. (1997). Monte Carlo studies of the dynamics of an interacting monodisperse magnetic-particle system. *Phys. Rev. B*, Vol. 56 (December, 1997), pp. 13983-13988, ISSN: 1098-0121
- Azeggagh, M. & Kachkachi, H. (2007). Effects of dipolar interactions on the zero-field-cooled magnetization of a nanoparticle assembly, *Phys. Rev. B*, Vol. 75 (May, 2007), pp. 174410 (1-9), ISSN: 1098-0121
- Blackwell, J. J.; Morales, M.P.; O'Grady, K.; Gonzalez, J.M.; Cebollada, F. & Alonso-Sanudo, M. (2003), Interactions and hysteresis behaviour of Fe/SiO<sub>2</sub> Nanocomposites, *J. Mag. Mag. Mat.*, Vol. 242-245 (2003), pp. 1103-1105, ISSN: 0304-8853.
- Bottoni, G.; Candolfo, D.; Cecchetti, M.; Corradi, A. R. & Masoli, F. (1993). Influence of packing density on the coercivity of iron particles for magnetic recording, *J. Mag. Mag. Mat.*, Vol. 120 (1993), pp. 167-171, ISSN: 0304-8853.
- Brown, W. F. Jr. (1963). Thermal Fluctuations of a Single-Domain Particle. *Phys. Rev.*, Vol.130 (June, 1963), pp. 1677-1686. ISSN: 0031-9007.
- Ceylan, A.; Bakker, C. C.; Hasanain, S. K. & Shah S. I. (2005). Nonmonotonic concentration dependence of magnetic response in Fe nanoparticle-polymer composites, *Phys. Rev. B*, Vol. 72, (October, 2005), pp. 134411, ISSN: 1098-0121
- Dennis, C. L.; Jackson, A. J.; Borchers, J. A.; Ivkov, R.; Foreman, A. R.; Lau, J. W.; Goernitz, E. & Gruettner, C. (2008). The influence of collective behavior on the magnetic and heating properties of iron oxide nanoparticles, *J. Appl. Phys.*, Vol. 103, (March, 2008), pp. 07A319 (1-3), ISSN: 0021-8979.
- Dormann, J. L.; Fiorani, D. & Tronc, E. (1999). On the models for interparticle interactions in nanoparticle assemblies: comparison with experimental results, *J. Mag. Mag. Mat.*, Vol. 202, (1999), 251-267, ISSN: 0304-8853.
- Fernández, J. F. & Alonso, J. J. (2009). Equilibrium spin-glass transition of magnetic dipoles with random anisotropy axes on a site-diluted lattice, *Phys. Rev. B*, Vol. 79, (June, 2009) 214424 (1-6), ISSN: 1098-0121.

- Friedman, J. R., Voskoboynik, U. & Sarachik, M. P. (1997). Anomalous magnetic relaxation in ferritin, *Phys. Rev. B*, Vol. 56 (November, 1997) pp. 56 10793, ISSN: 1098-0121.
- Garcia-Otero, J.; Porto, M.; Rivas, J. & Bunde, A. (2000). Influence of magnetic properties of ultra-fine ferromagnetic particles, *Phys. Rev. Lett.*, Vol. 84 (January, 2000), pp. 167-170, ISSN: 0031-9007.
- Georgescu, M.; Viota, J. L.; Klokkenburg, M.; Ern , B. H.; Vanmaekelbergh, D. & Zeijlmans van Emmichoven, P. A. (2006). Flux closure in two-dimensional magnetite nanoparticle assemblies, *Phys. Rev. B*, Vol. 73 (May, 2006), pp. 184415 (1-5), ISSN: 1098-0121.
- Georgescu, M.; Viota, J. L.; Klokkenburg, M.; Ern , B. H.; Vanmaekelbergh, D. & Zeijlmans van Emmichoven, P. A. (2008) Short-range magnetic order in two-dimensional cobalt-ferrite nanoparticle assemblies, *Phys. Rev. B*, Vol. 77 (January, 2008), pp. 024423 (1-6), ISSN: 1098-0121.
- Iglesias, O. & Labarta, A. (2004). Magnetic relaxation in terms of microscopic energy barriers in a model of dipolar interacting nanoparticles, *Phys. Rev. B*, Vol. 70, (February, 2004), 144401, ISSN: 1098-0121.
- Jonsson, T.; Mattsson, J.; Djurberg, C.; Khan, F. A.; Nordblad, P. & Svedlindh, P. (1995). Aging in a magnetic nanoparticle systems, *Phys. Rev. Lett.*, Vol. 75, (November, 1995), 4138, ISSN: 0031-9007.
- Kackachi, H.; Coffey, W. T.; Crothers, D. S. F.; Ezzir, A.; Kenedy, E. C.; Nogues, M. & Trone, E. (2000). Field dependence of the temperature at the peak of the zero-field-cooled magnetization. *J. Phys.: Condens. Matter*, Vol. 12, (November, 2000), pp. 3077, ISSN: 0953-8984.
- Kechrakos, D. & Trohidou, K. N. (1998). Magnetic properties of dipolar interacting single-domain particles, *Phys. Rev. B*, Vol. 58 (November, 1998), pp. 12169, ISSN: 1098-0121.
- Kechrakos, D. (2010). Magnetic nanoparticle assemblies. *Pre-print* (April, 2010).
- Kleemann, W.; Petravic, O.; Binek, Ch.; Kakazei, G. N.; Pogorelov, Y. G.; Sousa, J. B.; Cardoso, S. & Freitas, P. P. (2001). Interacting ferromagnetic nanoparticles in discontinuous  $\text{Co}_{80}\text{Fe}_{20} / \text{Al}_2\text{O}_3$  multilayers: From superspin glass to reentrant superferromagnetism. *Phys. Rev. B*, Vol. 63 (March, 2001), pp. 134423 (1-5), ISSN: 1098-0121.
- Knobel, M.; Nunes, W. C.; Socolovsky, L. M.; De Biasi, E.; Vargas, J. M. & Denardin, J. D. (2008). Superparamagnetism and other magnetic features in granular materials: A Review on Ideal and Real Systems, *J. Nanosci. Nanotech.*, Vol. 8 (2008), pp. 2836-2857, ISSN: 1550-7041.
- Lan, T. N. & Hai, T. H (2010). Monte Carlo simulation of magnetic nanoparticle systems. *Comput. Mater. Sci.*, Vol. 49 (March, 2010), pp. S287-S290, ISSN: 0927-0256.
- Lan, T. N. & Hai, T. H. (2010). Role of poly-dispersity and dipolar interaction in magnetic nanoparticle systems: Monte Carlo study. *J. Non-Cryst. Solids*. in press (December, 2010). ISSN: 0022 - 3093
- Luo, W.; Nagel, S. R.; Rosenbaum, T. F. & Rosensweig, R. E. (1991). Dipole interactions with random anisotropy in a frozen ferrofluid, *Phys. Rev. Lett.*, Vol. 67, (November, 1991), 2721, ISSN: 0031-9007.
- Malik, R.; Lamba, S.; Kotnala, R.K. & Annapoornil, S. (2010). Role of anisotropy and interactions in magnetic nanoparticle systems. *Eur. Phys. J. B*, Vol. 74, (February, 2010), pp. 75-80, ISSN: 1434-6028.

- Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H. & Teller, E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.*, Vol. 21 (March, 1953), pp. 1699-114 (1-6), ISSN: 0021-9606.
- Morup, S. & Tronc, E. (1994). Superparamagnetic relaxation of weakly interacting particles, *Phys. Rev. Lett.*, Vol. 72 (May, 1994), pp. 3278-3281, ISSN: 0031-9007.
- Neel, L. (1953). Thermoremanent magnetization of fine powders. *Rev. Mod. Phys.*, Vol. 25 (1953), pp. 293-295, ISSN: 0034-6861.
- Nunes, W. C.; Socolovsky, L. M.; Denardin, J.C.; Cebollada, F.; Brandl, A. L & Knobel, M. Role of magnetic interparticle coupling on the field dependence of the superparamagnetic relaxation time, *Phys. Rev. B*, Vol. 72 (December, 2005), pp. 212413 (1-4), ISSN: 1098-0121.
- Pankhurst, Q. A.; Connolly, J.; Jones, S. & Dobson, J. (2003). Applications of magnetic nanoparticles in biomedicine. *J. Phys. D: Appl. Phys.*, Vol. 36 (June, 2003), pp. R167 - R181, ISSN: 0022-3727.
- Parker, D.; Dupuis, V.; Ladieu, F.; Bouchaud, J.-P.; Dubois, E.; Perzynski, R. & Vincent, E. (2008), Spin glass behavior in an interacting  $\gamma$ -Fe<sub>2</sub>O<sub>3</sub> nanoparticle system, *Phys. Rev. B*, Vol. 77 (March, 2008), 104428(1-9), ISSN: 1098-0121.
- Perez, N.; Guardia, P.; Roca, A. G.; Morales, M. P.; Serna, C. J.; Iglesias, O.; Bartolomé, F.; García, L. M.; Batlle, X. & Labarta, A. (2008) Surface anisotropy broadening of the energy barrier distribution in magnetic nanoparticles, *Nanotechnology*, Vol. 19, (October, 2008), pp. 475704, ISSN: 0957-4484.
- Porto, M. (2005). Ordered systems of ultrafine ferromagnetic particles, *Eur. Phys. J. B*, Vol. 45 (January, 2005), pp. 369-375, ISSN: 1434-6028.
- Prokopenko, T. A.; Danilov, V. A.; Kantorovich, S. S. & Holm, C. (2009). Ground state structures in ferrofluid monolayers, *Phys. Rev. B*, Vol. 80 (September, 2009), pp. 031404 (1-13), ISSN: 1098-0121.
- Sappey, R.; Vincent, E.; Hadacek, N.; Chaput, F.; Boilot, J. P. & Zins, D. (1997). Nonmonotonic field dependence of the zero-field cooled magnetization peak in some systems of magnetic nanoparticles, *Phys. Rev. B*, Vol. 56 (December, 1997), 14551, ISSN: 1098-0121.
- Schaller, V.; Wahnström, G.; Sanz-Velasco, A.; Gustafsson, S.; Olsson, E.; Enoksson, & Johansson, C. (2009). Effective magnetic moment of magnetic multicore nanoparticles *Phys. Rev. B*, Vol. 80 (September, 2009), 092406, ISSN: 1098-0121.
- Serantes, D.; Baldomir, D., Pereiro, M., Arias, J.E.; Mateo-Mateo, C.; Bujan-Nunez, M.C.; Vazquez-Vazquez, C. & Rivas, J. (2008). Interplay between the magnetic field and the dipolar interaction on a magnetic nanoparticle system: A Monte Carlo study, *J. Non Crystalline Solids*, Vol. 354, (October, 2008), pp. 5224-5226, ISSN: 0022-3093.
- Tamion, A.; Raufast, C.; Hillenkamp, M.; Bonet, E.; Jouanguy, J.; Canut, B.; Bernstein, E.; Boisron, O.; Wernsdorfer, W. & Dupuis, V. (2010). Magnetic anisotropy of embedded Co nanoparticles: Influence of the surrounding matrix. *Phys. Rev. B*, Vol. 81 (April, 2010), pp. 144403 (1-6), ISSN: 1098-0121.
- Tan, R. P.; Lee, J. S.; Cho, J. U.; Noh, S. J.; Kim, D. K. & Kim, Y. K. Numerical simulations of collective magnetic properties and magnetoresistance in 2D ferromagnetic nanoparticle arrays, *J. Phys. D: Appl. Phys.*, Vol. 43, (March, 2010), pp. 165002 (1-8), ISSN: 0022-3727.

- Ulrich, M.; Garcia-Otero, J.; Rivas, J. & Bunde, A. (2003). Slow relaxation in ferromagnetic nanoparticles: Indication of spin-glass behavior, *Phys. Rev. B* Vol. 67 (January, 2003) 026616 (1-4), ISSN: 1098-0121.
- Verdes, C.; Ruiz-Diaz, B.; Thompson, S. M.; Chantrell, R. W. & Stancu, Al. (2002). Computational model of the magnetic and transport properties of interacting fine particles. *Phys. Rev. B* Vol. 65, (April, 2002), pp. 174417 (1-10), ISSN: 1098-0121.
- Walton, D. (2006). A theory for spin glass phenomena in interacting nanoparticle systems. *Pre-print*, (August, 2006).
- Yang, Y.; Shen, S.; Ye, Q.; Lin, L. & Huang, Z. (2006). The roles of the exchange and dipole couplings on the magneto-resistance for the nanoparticle arrays, *J. Mag. Mag. Mat.*, Vol. 303, (March, 2006), pp. e312-e314, ISSN: 0304-8853.
- Zheng, R. K.; Gu, H.; Xu, B. & Zhang, X. X. (2006). The origin of the non-monotonic field dependence of the blocking temperature in magnetic nanoparticles. *J. Phys.: Condens. Matter*, Vol. 18, (June, 2006), pp. 5905, ISSN: 0953-8984.
- Zheng, R. K.; Gu, H.; Zhang, B.; Liu, H.; Zhang, X. X. & Ringer, S. P. (2009). Extracting anisotropy energy barrier distributions of nanomagnetic systems from magnetization/susceptibility measurements. *J. Mag. Mag. Mat.*, Vol. 312 (2009), pp. L21-L27, ISSN: 0304-8853

# Monte Carlo Studies of Magnetic Nanoparticles

K. Trohidou and M. Vasilakaki

*Computational Materials Science Group, Institute of Materials Science,  
National Center of Scientific Research 'DEMOKRITOS',  
Patriarhou Grigoriou & Neapoleos St. 153-10 Agia Paraskevi, Athens,  
Greece*

## 1. Introduction

Magnetic nanoparticles are complex mesoscopic systems which have unique physical properties that clearly differ from those of atoms and bulk materials. They find numerous technological applications ranging from ultra-high-density recording media (Bader, 2006) to biomedicine (Pankhurst et al., 2003). The necessity to reduce the size of the nanoparticles for these applications have raised a key issue in their study which is their thermal stability.

The Monte Carlo (MC) simulation technique with the implementation of the Metropolis Algorithm (Metropolis et al., 1953) has been proved a very powerful tool for the systematic study of the magnetic behaviour of nanoparticles and nanoparticle assemblies. The two major advantages of this technique are a) the possibility for atomic scale treatment of the nanoparticles, so the details of their microstructure can be studied and b) the implementation of finite temperature through the Metropolis algorithm.

Although, the obtained dynamics in the Monte Carlo simulations is intrinsic and the time evolution of the system does not come from any deterministic equation for the magnetisation, the results of the Monte Carlo simulations reproduce qualitatively the trend of the experimental data (Binder 1987). Actually this good qualitative agreement between the simulation results and the experimental data enable us to have a better insight into the nanoscaled phenomena, though some of them stem from non-equilibrium processes (Landau & Binder, 2000).

A microscopic treatment of the magnetisation of ferromagnetic nanoparticles, using Monte Carlo techniques, was first developed by Binder and co-workers (Binder et al., 1970; Wildpaner 1974). An important demonstration of the work was the reduction of the magnetisation near the surface of the particle. Clearly this was to be expected because a surface spin has a smaller number of neighbours than it would have in bulk and, hence, experiences a reduced mean field. For very small particles (less than say 5 nm) the proportion of surface spins is such that they will make a major contribution to the magnetisation. As a result, the magnetisation will decrease with temperature over a range where the bulk magnetisation is roughly constant and deviations from Curie-law behaviour in the susceptibility are to be expected. In the period following the Monte Carlo work cited above, interest has been developed in finite-size scaling, and it is in this context that subsequent advances (Landau, 1976) in the nanoparticle magnetism have occurred. In addition, over the last decade there is a continuous effort to reduce the nanoparticles size and at the same time to overcome the thermal instability at room temperature (Skumryev et

al., 2003). This led to the study of composite nanoparticle magnetic structures with core/shell morphology.

In what follows we will review our MC simulation results for atomic scale modelling on spherical ferromagnetic (FM), antiferromagnetic (AFM), ferrimagnetic (FI) and composite nanoparticles with core/shell morphology. The magnetisation dependence on external parameters (temperature, applied field) and the intrinsic particle properties (size, size of shell and core, size and type of anisotropy, magnetic structure) are studied. Finite size effects and the role of the surface will be discussed for the FM, AFM and FI nanoparticles. In the case of the composite nanoparticles, which consist of a spherical ferromagnetic core surrounded by an antiferromagnetic (or ferrimagnetic) shell, we examine the effect of the interface between the ferromagnetic core and the shell of the particles, on their magnetic properties.

Finally our MC simulations results on the influence of the interparticle interactions on the macroscopic magnetic behaviour of assemblies of nanoparticles will be reviewed. The characteristics of the hysteresis loop and the temperature dependent magnetisation (Field Cooled (FC)/ Zero-Field Cooled (ZFC)) are studied numerically in magnetic nanoparticle assemblies using Monte Carlo simulations and the standard Metropolis algorithm. The computational technique for the calculation of the long ranged interparticle interactions will be discussed and results will be given for granular assemblies and ordered arrays of magnetic nanoparticles.

A discussion on potential applications and a comparison with experimental findings will be given in all cases.

## 2. Metropolis Monte Carlo simulation for the magnetic nanoparticles

The MC simulation technique is a standard method to study models of equilibrium or non equilibrium thermodynamic systems with many degrees of freedom by stochastic computer simulation. The starting point of the simulations is the appropriate choice of a model Hamiltonian and then the use of random numbers to simulate statistical fluctuations in order to generate the correct thermodynamical probability distribution according to a canonical ensemble (Binder 1986, 1987). In this way one may obtain microscopic information about complex systems which cannot be studied analytically or which might not be accessible in a real system. Contrary to Landau-Lifshitz or Langevin equations, Monte Carlo scheme provides the straightforward implementation of the temperature.

To simulate the magnetic nanoparticles and the nanoparticle assemblies and to derive thermodynamic averages, the elementary physical quantity that we use is the spin. In the case of the single nanoparticles we consider a classical spin at each atomic site and we simulate using the MC technique the stochastic movement of the system in the phase space. In the case of assemblies, we consider an effective spin to represent the magnetic state of each nanoparticle (Stoner & Wohlfarth, 1948).

The MC simulation consists of many elementary steps. In every elementary step a spin  $\hat{S}_{i_{old}}$  is randomly chosen from a system of  $N$  spins and an attempted new orientation  $\hat{S}_{i_{new}}$  of the spin is generated with a small random deviation  $\delta\hat{S}$ . The attempted direction is chosen in a spherical segment around the present orientation  $\hat{S}_{i_{old}}$ . Then the energy difference  $\Delta E$  between the attempted and the present orientation is calculated. In the Metropolis Monte Carlo algorithm, if  $\Delta E \leq 0$  the new orientation is accepted, if  $\Delta E > 0$  the attempted new orientation is accepted provided that a random number  $u$ , generated uniformly in the

interval (0,1), is less than the probability  $\exp(-\Delta E/k_B T)$ , otherwise the system remains to its present state (Binder, 1987). A complete MC step (MCS) consists of  $N$  elementary steps so that in any MC step on average every spin is considered once. With this algorithm, states are generated with a certain probability (Importance Sampling) and rejecting the first MCS that correspond to the thermalization process the desired average of a variable, namely the sum of the products of each value times the corresponded probability, simply become arithmetic average over the entire sample of states which is kept.

One common problem that appears during the MC simulation is that if we draw the attempted direction of every spin independently of the previous one, the system will always be superparamagnetic and no hysteresis will result, since it will be possible to explore the whole phase space independently of the temperature and due to the large fluctuations in every MCS it will escape very quickly from any metastable state responsible for hysteresis. By fixing to a certain limit the deviation  $\widehat{\delta S}$ , it is possible to modify the range of acceptance and model the real system more accurately (Garcia-Otero et al., 1999; Dimitrov & Wysin, 1996; Binder 1987) than choosing  $\hat{S}_{new}$  completely randomly and independently from  $\hat{S}_{old}$ . The MC acceptance rate can be set to some desired value (40-60%) (setting effectively the rate of motion in phase space). The use of such a kind of local dynamics permits to detect confinement in metastable states which are responsible for the hysteresis and to achieve true relaxation in different temperatures. Therefore we choose to perform the Metropolis MC in such a way that it samples the phase space "locally" with accepted ratio 50%.

Over the years several modified MC methods have been proposed to treat the problem of overcoming the local minima during the MC numerical procedure depending on the details of the system (e.g. Chantrell et al., 2001; Hinzke & Nowak, 1999; H. F. Du & A. Du, 2006). However the MC Metropolis algorithm works fast and efficiently in all cases.

In order to avoid trapping of the system at local minima, we start the numerical procedure from an unmagnetised sample at a high temperature above the critical temperature of the sample and we reduce the temperature gradually at a constant rate. At the temperatures above  $T_c$  we use more MC steps than at the lower temperature to let the system relax surpassing probable metastable states. Special care has been taken of the time and ensemble averaging of the magnetisation of the system by properly choosing the number of MC steps and a rather big number of different samples namely independent random number sequences corresponding to different realizations of thermal fluctuations.

The thermodynamic quantities that we calculate with the use of the MC Metropolis algorithm in the magnetic nanoparticles and their assemblies are the coercive field, the remanent magnetisation and the ZFC/FC magnetisation curves.

The coercive field ( $H_c$ ) are defined as the magnetic field required to reverse the magnetisation of the particle. In order to obtain the coercive field we calculate the complete hysteresis loop. A Field Cooling procedure is performed initially. Once the desired temperature is reached, we calculate the loop starting from the positive saturation and slowly decreasing the applied field in very small constant steps. In each value of the field, several MCS are executed, then the magnetisation is calculated and the field is changing again. We continue to reduce the field so the system goes to its negative saturation state. Then we increase again gradually the field until the system reaches its positive saturation. In this way a complete hysteresis loop is performed. The remanent magnetisation ( $M_r$ ) is taken at the zero field point of the descending magnetisation versus field curve.

The ZFC/FC magnetisation curves is obtained by the following steps: a) initially we start with the sample at very high temperature (above its critical temperature) and we

gradually reduce the temperature down to a very low value (close to zero), to obtain its ground state, b) at this very low temperature we apply a magnetic field and we start raising the temperature up to the maximum value that we had started, in this way we obtain the ZFC curve, c) finally in the presence of the magnetic field we reduce the temperature gradually down to the minimum value and in this way we obtain the FC magnetisation curve.

We have kept constant step rate of the magnetic field in the calculation of the hysteresis loops and of the temperature in the calculation of the ZFC/FC magnetisation curves (see e.g. Bahiana et al., 2004).

### 3. Model Hamiltonians for the magnetic nanoparticles and nanoparticle assemblies.

#### 3.1 Isolated magnetic nanoparticles

The Monte Carlo simulations are performed using the Metropolis algorithm as described in section 2. For the energy calculation of the single particle systems we use the following models:

a) In the case of FM, AFM and FI nanoparticles, we consider spherical nanoparticles with radius  $R$ , expressed in lattice spacings, on a simple cubic lattice. The outer layer of one lattice spacing is considered to be the surface of the nanoparticle in all cases. The spins in the particle interact with nearest neighbours Heisenberg exchange interaction, and at each crystal site they experience a uniaxial anisotropy. In the presence of an external magnetic field, the total energy of the system is:

$$E = -J \sum_{\langle i,j \rangle} \vec{S}_i \cdot \vec{S}_j - K_c \sum_{i \in \text{core}} (\vec{S}_i \cdot \hat{e}_i)^2 - K_s \sum_{i \in \text{surf}} (\vec{S}_i \cdot \hat{e}_i)^2 - \vec{H} \cdot \sum_i \vec{S}_i \quad (1)$$

Here  $S_i$  is the atomic spin at site  $i$  and  $\hat{e}_i$  is the unit vector in the direction of the easy axis at the site  $i$ . The first term gives the exchange interaction between the spins, the second is the anisotropy energy of the core, the third gives the anisotropy energy of the surface and the last term is the Zeeman energy. The core anisotropy ( $K_c$ ) is assumed uniaxial along the  $z$ -axis. The surface anisotropy ( $K_s$ ) is considered either radial or random. The exchange coupling constant  $J$  for the FM nanoparticles is taken equal to one, for the AFM ones  $-1$  and for the FI ones  $-1.5$ . The hysteresis loops are calculated after a Field Cooling procedure as described in Section 2.

b) For the composite spherical nanoparticles with FM core and an AFM shell or FI shell in a simple cubic lattice, we take into account explicitly the exchange interaction between the spins in the core, the interface, the shell and the surface considering also in this case nearest neighbours Heisenberg exchange interactions (Eftaxias & Trohidou, 2005; Eftaxias et al., 2007; Vasilakaki & Trohidou, 2009). The energy of the system is given as:

$$E = -J_{\text{FM}} \sum_{\langle i,j \in \text{FM} \rangle} \vec{S}_i \cdot \vec{S}_j - J_{\text{SH}} \sum_{\langle i,j \in \text{SH} \rangle} \vec{S}_i \cdot \vec{S}_j - J_{\text{IF}} \sum_{\langle i,j \in \text{IF} \rangle} \vec{S}_i \cdot \vec{S}_j - \sum_{i \in \text{FM}} K_{\text{IFM}} (\vec{S}_i \cdot \hat{e}_i)^2 - \sum_{i \in \text{SH}} K_{\text{ISH}} (\vec{S}_i \cdot \hat{e}_i)^2 - \vec{H} \cdot \sum_i \vec{S}_i \quad (2)$$

The first, second and third terms give the core, shell and interface exchange interaction respectively. The exchange coupling constant  $J_{\text{FM}}$  for the core spins is taken equal to one. We

set the exchange coupling constant of the  $J_{SH} = -J_{FM}/2$ . The exchange coupling constant of the interface  $J_{IF}$  is equal to  $J_{SH}$  in size and the interaction is taken ferromagnetic. The fourth term gives the anisotropy energy of the FM core. If the site  $i$  lies in the outer layer of the FM core, the anisotropy is defined as  $K_{iFM}=K_{IF}$  and  $K_{iFM}=K_C$  elsewhere. The anisotropy always is considered uniaxial along the  $z$ -axis in the core and along the interface. The fifth term gives the anisotropy energy of the AFM shell and it is considered either along the  $z$ -axis or random. If  $i$  lies in the outer layer of the shell then the anisotropy is defined as  $K_{iSH} = K_S$  and  $K_{iSH} = K_{SH}$  elsewhere. The surface anisotropy is taken as random. So  $K_C, K_{IF}, K_{SH}, K_S$ , denote the core, the interface, the shell and the surface anisotropy respectively. The last term gives the Zeeman energy.

We simulate a Field Cooling procedure starting at a temperature which is between the Curie temperature  $T_c$  of the ferromagnetic core and the critical temperature of the antiferromagnetic or ferrimagnetic shell; consequently we cool the nanoparticle at a constant rate in the presence of a magnetic field  $H_{cool}$  along the  $z$ -axis. The resulting hysteresis loops have a horizontal and vertical asymmetry. The value of the loop shift along the field axis is expressed by the exchange bias field  $H_{ex} = -(H_{right}+H_{left})/2$ , and the coercive field is defined as  $H_c = (H_{right}-H_{left})/2$ ,  $H_{right}$  and  $H_{left}$  being the points where the loop intersects the field axis. The vertical shift (DM) that expresses the asymmetry along the magnetisation axis, is given as  $DM = (M_{up}-M_{down})/2$ ,  $M_{up}$  and  $M_{down}$  being the points where the loop intersects the  $M$ -axis.  $M_r$  is normalized to the magnetisation at saturation ( $M_s$ ).

### 3.2 Assemblies of magnetic nanoparticles

We considered two types of nanoparticle assemblies: three-dimensional (3D) randomly placed magnetic nanoparticle assemblies and quasi two-dimensional (2D) ordered arrays of magnetic nanoparticles.

In the first type we consider spherical nanoparticles with diameter  $D$  located randomly inside a cubic box with edge length equal to  $L$ . The particle assembly is assumed monodisperse. To avoid the overlap problem, the space inside the box is discretised by a simple (or face centered) cubic lattice with lattice constant equal to the particle diameter. The magnetic state of each particle is described, according to the Stoner-Wohlfarth model (Stoner & Wohlfarth, 1948), by a classical spin vector ( $\hat{S}_i$ ) with an anisotropy axis in a random direction ( $\hat{e}_i$ ). The particles interact via long range dipolar forces and via exchange forces, when they are sufficiently close. The total energy of the assembly is given by equation 3:

$$E = g \sum_{i,j} \frac{(\hat{S}_i \cdot \hat{S}_j) - 3(\hat{S}_i \cdot \hat{R}_{ij})(\hat{S}_j \cdot \hat{R}_{ij})}{R_{ij}^3} - J \sum_{\langle i,j \rangle} \hat{S}_i \cdot \hat{S}_j - k \sum_i (\hat{S}_i \cdot \hat{e}_i)^2 - h \sum_i (\hat{S}_i \cdot \hat{H}) \quad (3)$$

where  $\hat{S}_i$  is the direction of the spin of the nanoparticle  $i$ ,  $\hat{e}_i$  is the easy axis direction,  $R_{ij}$  is the centre-to-centre distance between particles  $i$  and  $j$ , measured in units of the particle diameter and hats indicate unit vectors. The first term gives the dipolar energy where  $g$  is the dipolar strength defined as  $g = \mu^2/D^3$ . The second term gives the exchange energy with exchange strength  $J$ . The third term gives the anisotropy energy with the anisotropy constant defined as  $k = K_1 V_0$  and the last term gives the Zeeman energy with  $h = \mu H$  where  $\mu = M_s V_0$  is the nanoparticle magnetic moment and  $V_0$  the nanoparticle volume. The exchange coupling exists only between particles in contact (nearest neighbours).

The particles occupy at random the sites within the cube with an occupation probability  $p$ . The MC cell is repeated periodically and the Ewald summation technique is implemented to calculate the long range part of the dipolar energy (Kretschmer & Binder, 1979). Satisfactory convergence with the Ewald technique is obtained using repetitions of the central Monte Carlo cell along each of the three Cartesian axes (Kechrakos & Trohidou, 1998). As a test of convergence of the Ewald series we calculated the value of the local field at a site of a fully aligned ferromagnetic state of a crystalline ( $p=1$ ) assembly. The theoretical value  $H_0 = (4\pi/3) M_s$  was reproduced with accuracy  $10^{-4}$ .

For the model of the quasi 2D nanoparticle ordered arrays we consider identical spherical particles with diameter  $D$  forming a two-dimensional triangular lattice in the  $xy$ -plane and lattice constant  $d \geq D$ . We construct a nanoparticle-assembled film with finite thickness (1-4 monolayers (ML)) by placing particles in the upper layer above alternate interstices in the lower one (Puntes et al., 2001). Structural defects are considered only in the uppermost ML. The nanoparticles are single-domain, they possess uniaxial anisotropy in a random direction and they interact via dipolar forces. The total energy of the system is given again by equation (3); in this case the exchange interaction term is not taken into account, since in the ordered arrays the nanoparticles are not in contact (Kechrakos & Trohidou, 2002). We used periodic boundaries in the  $xy$ -plane and free boundaries in the  $z$ -axis. The dipolar interactions were treated without truncation using the Ewald summation method for a quasi-two-dimensional system (Grzybowski et al., 2000).

In all cases of nanoparticles and nanoparticle systems, the fields  $H_{cool}$  and  $H_c$  are given in units of  $J/g\mu_B$ , the temperature  $T$  in units  $J/k_B$ , and the anisotropy coupling constants  $K$  in  $J$ . In our simulation we have used from 18.000 up to 40.000 Monte Carlo steps per spin (Binder 1987), depending on the system size. The results are averaged over 10-30 samples with different spin configurations for the single particles and the assemblies and in the case of the random assemblies different spatial configurations for the nanoparticles have also been considered.

#### **4. MC simulation results for isolated FM, AFM and FI nanoparticles.**

Surface effects, resulting basically from the symmetry breaking of the lattice, and finite-size effects have a strong influence on the magnetic properties of single-domain nanoparticles. As the particle size decreases in FM, AFM and FI nanoparticles, new or modified magnetic properties come to light opening new horizons for research, production of novel nanostructures and technological applications. A common feature to the study of single domain nanoparticles has been the large deviations from the uniform reversal magnetisation of the Stoner-Wohlfarth model and the existence of a progressive switching of the spins that caused the need for a different numerical treatment including micromagnetic details describing the surface effects.

Our Monte Carlo simulations have proved to be efficient to examine the influence of the surface spins and to incorporate the surface effects to the magnetic behaviour of ferromagnetic, antiferromagnetic and ferrimagnetic nanoparticles with uniaxial, random and radial surface anisotropy.

##### **4.1 Ferromagnetic nanoparticles**

FM nanoparticles show enhanced magnetic moments and enhanced effective magnetic anisotropy values as the size decreases. This has been associated with the influence of the

surface atoms that become more significant with decreasing size due to increasing surface to volume ratio (Chen et al., 1998; Respaud et al., 1998; Bødker et al., 1994; Jamet et al., 2001). Also a decrease of the net surface magnetisation has been attributed to the effect of increasing surface disorder or surface spin canting or even to the existence of a dead magnetic layer (Curiale et al., 2009).

We have calculated the magnetisation of FM particles by the MC simulation technique using equation (1) for the energy minimisation. The results in the absence of any surface anisotropy are shown in figure 1(a) for two spherical nanoparticles. In this figure the temperature dependence of the magnetisation is shown for particles with uniform uniaxial anisotropy  $K_c$  with the easy axis along the  $z$ -direction,  $K=K_c=K_s=0.1J$  (full symbols). For these ferromagnetic particles, the decrease in the magnetisation with decreasing  $R$  is well known and it is ascribed to the increasing role played by the surface as  $R$  becomes smaller (Trohidou et al., 1998 a, 1998 b; Gangopadhyay et al., 1992; Dimitrov & Wysin, 1994). The effect of introducing radial surface anisotropy is next considered, for these two particles (open symbols). The surface anisotropy  $K_s$  is one order of magnitude higher than the core  $K_c=0.1J$ ,  $K_s=10K_c$  with the easy axis normal to the surface at each site. We observe in this case a reduction of the magnetisation due to the surface disorder introduced by the radial anisotropy and a more rapid fall of the magnetisation with temperature. The radial direction of the easy axis orientation of the strong surface anisotropy when averaged over the whole surface of the spherical particle tends to eliminate the contribution to the magnetisation from the surface layer. This behaviour is in agreement with experimental findings on  $(Fe_{0.26}Ni_{0.74})_{50}B_{50}$  nanoparticles (De Biasi et al., 2002), on metallic Fe nanoparticles (Bodker et al., 1994) and Fe nanoparticles (Chen et al., 1998).

In figure 1(b) the coercive field versus temperature for these two nanoparticles is displayed. Full symbols represent results for uniform  $z$ -axis anisotropy  $K=K_c=K_s=0.1J$  and open symbols for  $z$ -axis core anisotropy  $K_c=0.1J$  and radial surface anisotropy of size  $K_s=10K_c$ . The observed behaviour for the particles with uniform anisotropy is the predicted one from the phenomenological model of Kneller and Luborsky (Kneller & Luborsky, 1963). The bigger particle has the higher coercivity and this behaviour is valid for all temperatures.

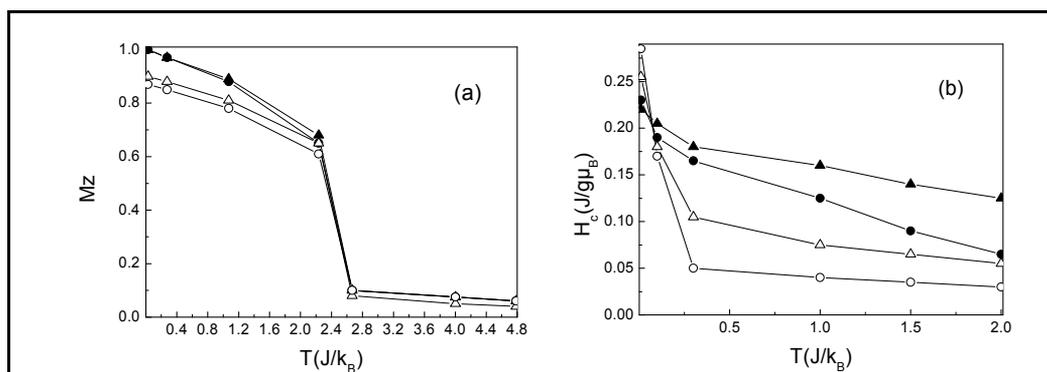


Fig. 1. Magnetisation versus temperature (a) and  $H_c$  versus temperature (b) for ferromagnetic particles with sizes:  $R=8.0$  (circles);  $R=12.0$  (triangles) with uniform uniaxial anisotropy  $K=0.1J$  (full symbols) and with radial surface anisotropy  $K_s=10K_c$  (open symbols).

The behaviour of the particles with the radial surface anisotropy is quite different. Increase of the low temperature coercivity for the particles with radius  $R=8$  is observed and an inversion with the size. As the temperature increases this behaviour is reversed, also we observe a steeper drop with temperature than in the uniform anisotropy case in agreement with Garanin & Kachkachi, 2003. This is due to the fact that the thermal fluctuations mask the surface contribution.

#### 4.2 Antiferromagnetic nanoparticles

AFM nanoparticles have attracted major interest since the pioneering study of Néel (Néel, 1953). The imbalance of the spin population on the antiparallel sublattices (uncompensated spins denoted by  $N_u$ ) gives a finite moment to the nanoparticles. Néel (Néel, 1953) first pointed out that the uncompensated spins of the AFM nanoparticles are lying on the surface (Trohidou, 2005). Experimental findings on AFM nanoparticle systems showed deviations of the magnetisation curves from the Langevin function above the blocking temperature, low blocking temperatures, shifted loops and high coercivities in the low temperature regime (Kodama R H 1999; Mørup & Hansen 2005). These findings indicate the interplay between size and surface effects and they can be described by a core / shell model where the nanoparticle consists of an antiferromagnetically ordered core and a disordered surface shell. This shell represents the frustrated magnetic state at the surface (Winkler et al., 2005; Bhowmik et al., 2004).

We have simulated four AFM spherical nanoparticles with very similar sizes but different numbers of uncompensated spins ( $R=7.5$  lattice spacings  $N=1791$  and  $N_u=79$ ,  $R=7.75$  with  $N=1935$  and  $N_u=17$ ,  $R=8.1$  with  $N=2205$  and  $N_u=83$  and  $R=8.5$  with  $N=2553$  and  $N_u=25$ ),  $N$  is the total number of spins in the nanoparticle. We introduce uniaxial anisotropy for the core spins and a strong random anisotropy at the surface to simulate the experimentally observed spin-glass like phase. The parameters for the anisotropy strength are  $K_c=0.1J$  in the core and  $K_s=1.5J$  at the surface (Vasilakaki & Trohidou, 2008).

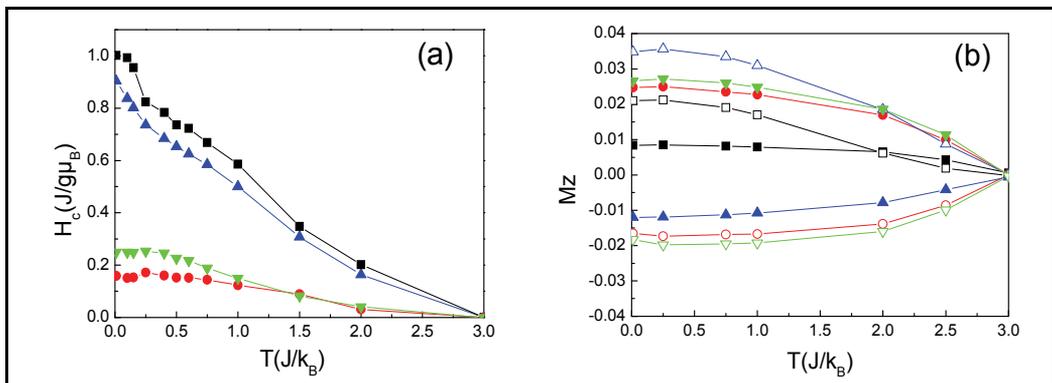


Fig. 2. Temperature dependence (a) of the coercive field ( $H_c$ ) and (b) of the z component of the magnetisation ( $M_z$ ) in the core (full symbols) and at the surface (open symbols) for AFM nanoparticles of radii  $R=7.5$  (squares),  $R=7.75$  (circles),  $R=8.1$  (up triangles) and  $R=8.5$  (down triangles) lattice spacings with random surface anisotropy.

By implementing the Monte Carlo simulation technique and the Metropolis algorithm for the energy minimization (given in eq. 1), we calculated the coercive field and the magnetisation versus temperature (figure 2). We observe that the  $H_c(T)$  curves are very close

for the nanoparticles with similar size of  $N_u$ . Also it appears that at intermediate temperatures they follow the  $N_u/N$  scaling law as discussed in (Néel, 1953; Trohidou, 2005). In figure 2(b) we give the temperature dependence of the core (full symbols) and the surface (open symbols) contributions to the  $z$  component of the magnetisation for the four nanoparticles. It can be seen clearly that the coercive field follows the surface behaviour.

For comparison in figure 3 we present results for the same nanoparticles for radial surface anisotropy and the same anisotropy strengths as in the random surface anisotropy case. The most striking feature in this figure is the appearance of a peak of the  $H_c$  at the same temperature roughly for all the nanoparticles. From figure 3(b) we can see that this is the temperature where the surface spins have a peak in their magnetisation so the surface spins drag the core ones and this accounts for the peak in the coercive field. Then as the temperature increases the surface spins move due to thermal fluctuations causing the gradual decrease of magnetisation. This behaviour is in agreement with that of layered systems with competing interactions (Leighton et al., 2002). It is apparent from figure 2(a) that also in this case of anisotropy the  $H_c$  scales with the  $N_u/N$  ratio at intermediate temperatures.

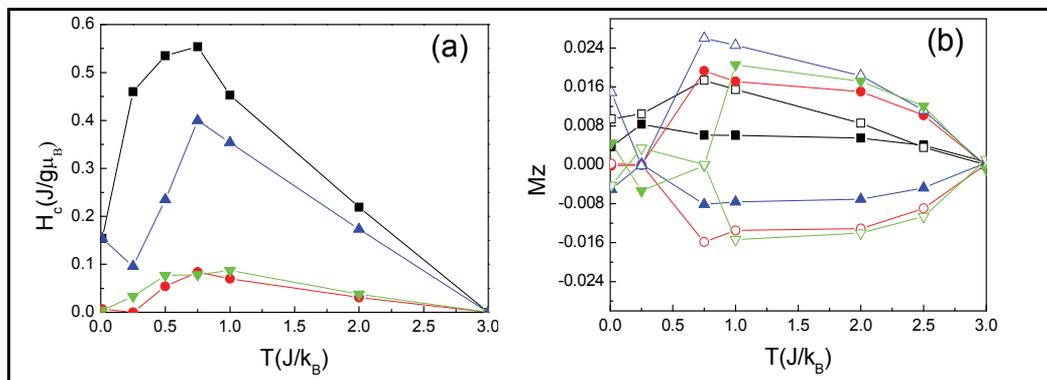


Fig. 3. Temperature dependence (a) of the coercive field ( $H_c$ ) and (b) of the  $z$  component of magnetisation ( $M_z$ ) in the core (full symbols) and at the surface (open symbols), for AFM nanoparticles of radii  $R=7.5$  (squares),  $R=7.75$  (circles),  $R=8.1$  (up triangles) and  $R=8.5$  (down triangles) lattice spacings with radial surface anisotropy.

The negative surface contribution in figures 2(b) and 3(b) from the surface magnetisation in the nanoparticle with radii  $R=7.75$  and  $8.5$  lattice spacings and very small number of uncompensated spins ( $N_u=17, 25$ ) is due to the non-uniform distribution of the small number of surface spins as discussed in (Trohidou et al., 1998). Also we observe in all nanoparticles a small increase of the core magnetisation in both cases of anisotropy at finite temperatures due to the fact that as temperature increases we have contribution from sublattice spins in the magnetisation as it has been discussed in (Mørup et al., 2007; Brown et al., 2005).

We also examined how the different surface anisotropy strengths modify the coercive field behaviour as a function of temperature (Vasilakaki & Trohidou, 2008). Our simulations demonstrated that the  $H_c(T)$  behaviour depends on the relative  $K_s/K_c$  ratio. The decreasing surface anisotropy allows the core spins to drag continuously the surface ones as they become disordered with increasing temperature. Our results are in agreement with the experimental findings of (Winkler et al., 2008) on noninteracting NiO nanoparticles.

### 4.3 Ferrimagnetic nanoparticles

In the case of ferrimagnetic nanoparticles also the surface role becomes important with the decrease of the size (Iglesias & Labarta, 2004; Leite et al., 2005; Martinez et al., 1998) and even larger than the antiferromagnetic surface because the net magnetic component of the ferrimagnetic surface is larger. The idea also of a noncollinear spin arrangement at the surface responsible for the moment reduction was proposed very early (Coeey, 1971).

In the case of ferrimagnetic nanoparticles we discuss the effect of the uncompensated spins on their coercive behaviour. We use the same parameters for the particle sizes, the core and the surface anisotropy as in the AFM nanoparticles for comparison.

In figure 4(a) we give the results for the coercive field versus temperature and in figure 4(b) we have plotted the temperature dependence of core and the surface contributions to the z component of the magnetisation for the four FI nanoparticles for random surface anisotropy and in figures 5(a) and 5(b) for radial surface contribution. At temperatures close to  $T=0$   $J/k_B$ , it is the core that contributes to the coercive behaviour of the biggest particles and we have an almost identical contribution of the surface for the two smaller ones. As a result all four nanoparticles have the same coercive field. As the temperature increases we have some contribution from the surface of the bigger nanoparticle and this gives a slower decrease of the coercive field with temperature for the bigger ones. This can be clearly seen from the magnetisation behaviour of the core and the surface in figure 4(b). The larger nanoparticles at very low temperature have negligible surface contribution to the magnetisation in comparison to the smaller ones. As the temperature increases the competition between the spin canting and the thermal fluctuations causes an increase in the surface contribution to the magnetisation for these nanoparticles. The details of this surface contribution depend on the distribution of the uncompensated spins on the surface of the particles. The magnetisation behaviour observed in figure 4(b) is in agreement with the magnetisation behaviour of  $Fe_2O_3$  nanoparticles in (Kachkachi et al., 2000).

The dependence of the magnetisation and the coercive field with temperature does not change when we replace the random anisotropy with radial at the surface confirming that here the magnetic behaviour is less sensitive to the surface contribution as it can be seen from figures 5.

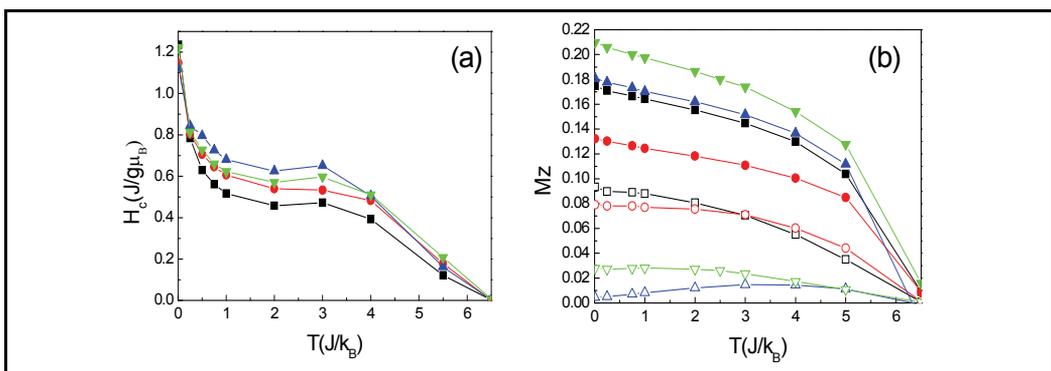


Fig. 4. Temperature dependence (a) of the coercive field ( $H_c$ ) and (b) of the z component of the magnetisation ( $M_z$ ) in the core (full symbols) and at the surface (open symbols) for FI nanoparticles of radii  $R=7.5$  (squares),  $R=7.75$  (circles),  $R=8.1$  (up triangles) and  $R=8.5$  (down triangles) lattice spacings with random surface anisotropy.

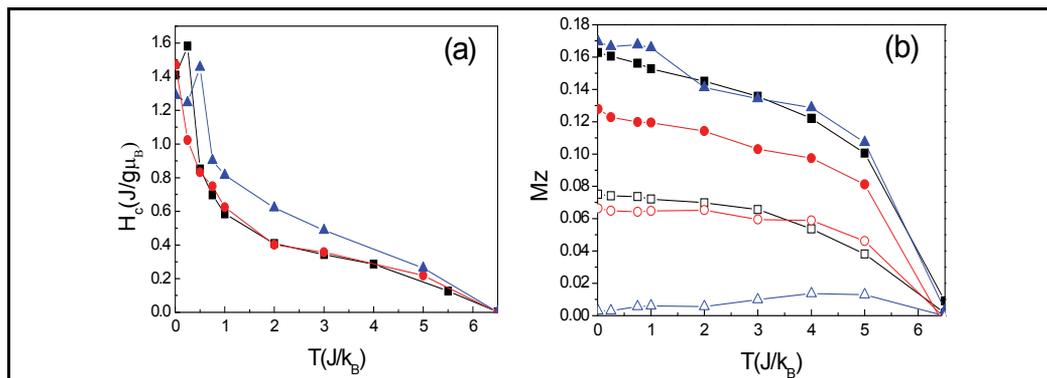


Fig. 5. Temperature dependence (a) of the coercive field ( $H_c$ ) and (b) of the z component of the magnetisation ( $M_z$ ) in the core (full symbols) and at the surface (open symbols) for FI nanoparticles of radii  $R=7.5$  (squares),  $R=7.75$  (circles) and  $R=8.1$  (up triangles) lattice spacings with radial surface anisotropy.

## 5. MC results for composite magnetic nanoparticles with FM core / AFM or FI shell morphology.

Fine particles with core/shell morphology have attracted great research interest due to rich and often unusual magnetic properties (Nogues et al., 2005). They exhibit enhancement of the coercive field and thermal stability at very small sizes, properties desirable for permanent magnets and the high density magnetic recording materials.

These properties are observed in all magnetic nanostructured materials with two different spin structures in contact. They exhibit asymmetry, on the magnetic field axis, of their hysteresis loops which is caused by a unidirectional anisotropy, the exchange anisotropy, induced by the exchange coupling at the interface between the different spin structures, when they are cooled down in a static magnetic field ( $H_{cool}$ ). The effect was first observed in field cooled Co/CoO composite nanoparticles with ferromagnetic core and antiferromagnetic (AFM) shell morphology (Meiklejohn & Bean, 1957) and it is known as the exchange bias effect. Although in the decades that followed the exchange bias research focused mainly on thin film systems, the production of magnetic nanoparticles with core/shell morphology and their study have renewed interest in the exchange bias phenomena (Nogues et al., 2005).

In addition to the exchange bias shift a vertical shift (DM) of the hysteresis loops has also been observed in FM/AFM core/shell nanoparticles attributed to the uncompensated spins of the shell (Passamani, 2006).

We have investigated the exchange bias mechanism and the factors that influence the exchange bias behaviour in FM core/AFM shell nanoparticles. In our studies (Zianni & Trohidou, 1998; Eftaxias et al., 2007) on FM core/AFM shell nanoparticles, we used the MC simulation technique employing the Metropolis algorithm to study the factors that influence their exchange bias behaviour. We found (Eftaxias & Trohidou, 2005; Eftaxias et al., 2007) that the exchange bias field at very low temperature is approximately constant after the second AFM layer, a result which is in agreement with the experimental findings of (Morel et al., 2004) where a very fast stabilization of the exchange bias field with oxygen dose in Co/CoO nanoparticles is observed and in Co/CoO nanoparticles embedded in an  $Al_2O_3$

matrix (Nogues et al., 2006). As the temperature increases, more AFM layers are needed to increase and stabilize the exchange bias field, because the thermal fluctuations mask the interface contribution, and therefore a thicker shell is required to stabilize the interface contribution; and after a certain number of AFM layers, roughly when the shell size becomes initially equal in size to the core and then further increases, the exchange bias field is decreasing because of the enhancement of the AFM contribution that masks the interface role.

We next studied isolated composite nanoparticles with a FM core and FI disordered shell morphology, in order to investigate the underlying mechanism for the exchange bias effects in these systems. The anisotropy constant for the core is  $K_c=0.05 J_{FM}$ , for the ferromagnetic interface  $K_{IF/FM} = 0.5 J_{FM}$  one order of magnitude larger than  $K_c$ , for the ferrimagnetic interface, the shell and the surface:  $K_{IF/FI} = 1.5 J_{FM}$ ,  $K_{SH} = 1.5 J_{FM}$  and  $K_s = 1.5 J_{FM}$  respectively. We introduce the strong random anisotropy in the shell and at surface in order to simulate a spin-glass like phase.

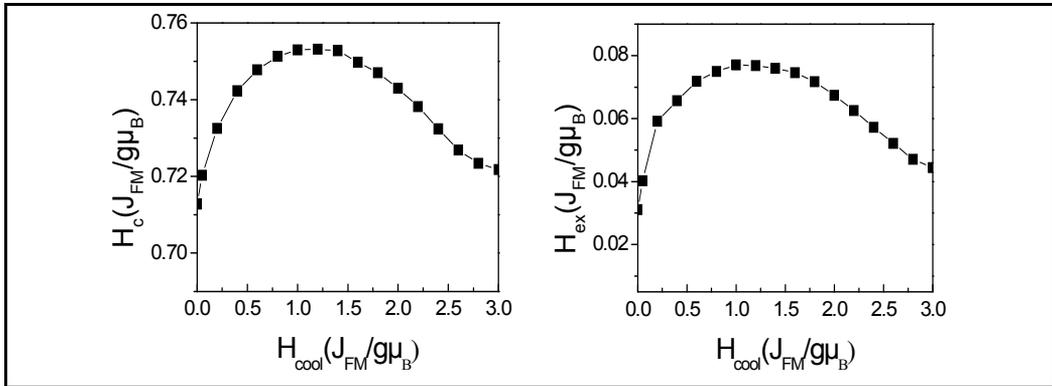


Fig. 6. Cooling field dependence of the coercive field ( $H_c$ ), and the exchange bias field ( $H_{ex}$ ) for a nanoparticle with core size 5 lattice spacings and shell thickness 7 lattice spacings.

The cooling field dependence of the  $H_{ex}$ ,  $H_c$  and  $M_r$  is given in figure 6 as a function of the applied cooling field  $H_{cool}$  for a nanoparticle with FM core 5 lattice spacings and FI shell 7 lattice spacings. Initially as the  $H_{cool}$  increases, it causes an increase in both  $H_{ex}$  and  $H_c$ . The gradual increase of  $H_{cool}$  tends to align a certain amount of FI spins at the interface along the field direction. After some  $H_{cool}$  value further increase in the cooling field, results in a decrease of these two quantities. For these higher cooling field values the Zeeman coupling between the field and the FI spins dominates the magnetic interactions inside the system. So the FI spins follow the applied field and as a result the exchange bias field and the coercive field decrease. Our MC results reproduced very well the behaviour of the cooling field dependence of  $H_c$ ,  $H_{ex}$  and  $M_r$  of Fe/FeO nanoparticles systems in (Baker et al, 2004; Del Bianco et al, 2004).

The influence of the shell thickness in the behaviour of the hysteresis loop, starting the field cooling procedure with a field  $H_{cool}=0.4 J/g\mu_B$  in nanoparticles with FM core/FI shell is discussed. We consider four particles with total radii  $R=9.0$ ,  $R=12.0$ ,  $R=14.0$  and  $R=20.0$  lattice spacings. They all have the same core size of 5 lattice spacings and FI shell thickness 4, 7, 9 and 15 lattice spacings respectively. The surface thickness is one lattice spacing, in all cases. The results for the hysteresis loops for these four particles are shown in figure 7(a) at a

very low temperature  $T=0.01J_{\text{FM}}/k_B$ . As we can see the hysteresis loops are shifted and the nanoparticles with the bigger shell thickness have the bigger shift and the bigger coercive field. So both  $H_c$  and the  $H_{\text{ex}}$  increase with the shell thickness while  $M_r$  decreases. This is in agreement with the experimental findings of Baker and his collaborators on Fe/FeO nanoparticles (Baker et al., 2004). We also observe a small vertical shift in the hysteresis loops of the two nanoparticles with the smaller shell thickness. The asymmetry in the magnetisation axis disappears for the two bigger nanoparticles. The hysteresis loop for the nanoparticle with the lower shell thickness has a shoulder, characteristic of a two-phase system. This shoulder disappears as the shell thickness increases, because the shell dominates in the hysteresis behavior of the sample. In figure 7(a) we observe that the hysteresis loops of the nanoparticles with the bigger shell thickness are less saturated, due to the enhancement of the spin-glass like behaviour (Binder & Young, 1986).

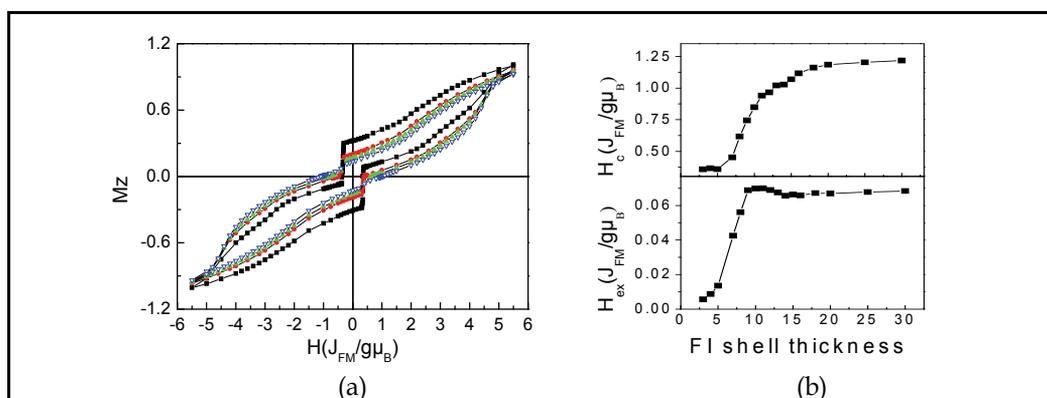


Fig. 7. (a) Hysteresis loops of core/shell nanoparticles with core radius  $R_c=5$  lattice spacings and shell thickness 4 (squares), 7 (circles), 9 (up triangles), 15 (down triangles) lattice spacings respectively (b) Shell thickness dependence of the coercive field ( $H_c$ ), the exchange bias field ( $H_{\text{ex}}$ ).

In figure 7(b)  $H_c$  and  $H_{\text{ex}}$  have been plotted as a function of the shell thickness. We observe that: a) the coercive field increases continuously with the shell thickness, b) the  $H_{\text{ex}}$  increases slowly with the increase of the shell thickness and then it remains constant.

The increase of  $H_c$  with the increase of the disordered layer is expected due to the fact that the thicker shell has bigger number of disordered spins. Our MC results agree with the experimental findings for Fe/Fe oxide core/shell nanoparticles by (Baker et al., 2004). In experimental studies of FM/AFM (Nogues & Schuller, 1999) or FM/FI bilayers (Lin et al., 1994) and Co/CoO (Peng et al., 2000) nanoparticles, they found that  $H_{\text{ex}}$  increases and then saturates for large AFM thickness. In the case of the ferrimagnetic shell nanoparticles it needs more layers even at low temperature in comparison with the AFM shell case (Eftaxias et al., 2007) for the appearance of the exchange bias effects.

The core size dependence of the exchange bias field and the coercive field is next considered. In figure 8 we present results for the  $H_c$  and  $H_{\text{ex}}$  versus temperature for three nanoparticles with the same shell thickness of 7 lattice spacings and three different core sizes of 3, 5 and 10 lattice spacings. As we can see, at very low temperature smaller core radius have the bigger coercive and exchange bias field, in agreement with the experimental findings of (Del Bianco et al., 2003). This is due to the fact that the biggest contribution from

the interface is obtained in the nanoparticle with the smallest core radius. The coercive field is decreasing faster with temperature in the case of the smaller core nanoparticles than in the bigger ones. There is a crossing temperature above which the behaviour is changed and the nanoparticles with the bigger core radius have higher coercivity. This is the temperature at which the shell becomes totally disordered. Above this temperature the coercivity follows the temperature dependence of the core. The smaller in size core becomes faster superparamagnetic. This is in agreement with experimental findings by (S. Gangopadhyay et al., 1992) on Fe nanoparticles surrounded by a disordered iron oxide shell. In the case of the two smaller nanoparticles the  $H_c$  decays exponentially as in the case of the varying shell thickness (see Fig. 7) due to the dominance of the disordered shell. For the biggest nanoparticle  $H_c$  has a monotonic temperature dependence due to the dominant ferromagnetic character. The exchange bias field vanishes very quickly with the increase of the temperature.

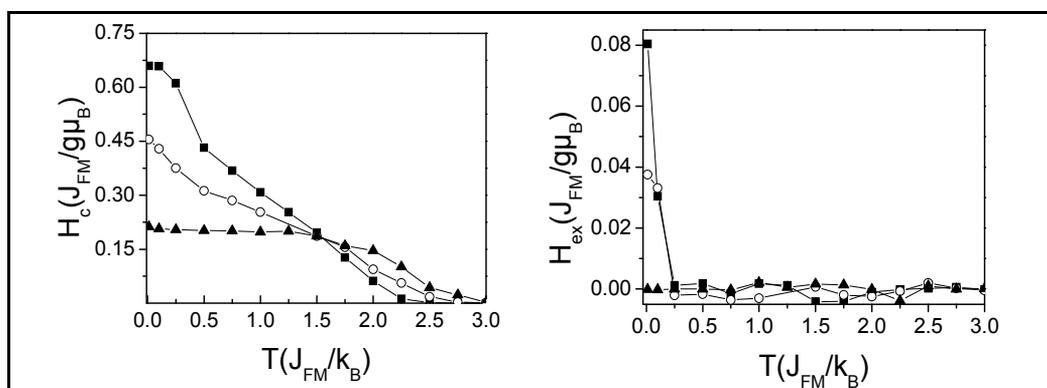


Fig. 8.  $H_c$  and  $H_{ex}$  as a function of temperature for nanoparticles with shell thickness 7 lattice spacings and core 3(squares), 5(circles) and 10(triangles) lattice spacings respectively.

For the case of composite nanoparticles with spin glass-like shell the aging and training effect on the  $H_c$  and  $H_{ex}$  have been also studied. These effects are present in the FM core /FI spin glass-like shell systems, since one of the characteristics of spin glass systems is their multiple energy configuration of the ground state (Binder & Young, 1986). So the frozen spins, which are originally aligned in the cooling field direction, may change their directions and fall into other metastable configurations during the hysteresis measurements. This characteristic of spin-glass-like phase essentially influences the exchange bias behaviour of the system and results in a decrease of  $H_c$  and  $H_{ex}$  with the field cycling. The behaviour of the training effect for the composite FM core/FI spin glass-like shell nanoparticles will depend on its microstructure characteristics. In figure 9 we have plotted the  $H_{ex}$  and  $H_c$  as a function of the loop cycling for three nanoparticles with core radius 5 lattice spacings and shell thickness 7, 9 and 15 lattice spacings (Vasilakaki & Trohidou, 2009). After a field cooling process, the hysteresis loop was calculated six consecutive times at temperature  $T=0.01 J_{FM}/k_B$ . We observe that  $H_{ex}$  in the case of the nanoparticle with shell thickness 7 lattice spacings after the first loop has a big reduction, while  $H_{ex}$  for the other two nanoparticles has a small reduction with the loop cycling and very similar. These two nanoparticles have the same size of the exchange bias field as it can be seen from figure 6.  $H_c$  has a bigger reduction with the loop cycling for the smaller shell nanoparticle than in the other two. This behaviour indicates that the interface has the major contribution in the

training effect, for the chosen nanoparticles. As the shell thickness decreases we expect contribution from the shell and the core to the training effect. So a  $\sim 50\%$  reduction of  $H_{\text{ex}}$  during cycling has been observed in FM/FI nanoparticles (Peng et al., 2002; Trohidou et al., 2007) with small shell thickness. Whereas a  $\sim 12\%$  reduction in the case of Co/CoO nanoparticles (Peng et al, 2000). In any case we expect that the behaviour of the training effect depends not only on the magnetic microstructure but on other factors, in agreement with the experimental observations in FM/AFM bilayer systems (Nogués et al., 2005).

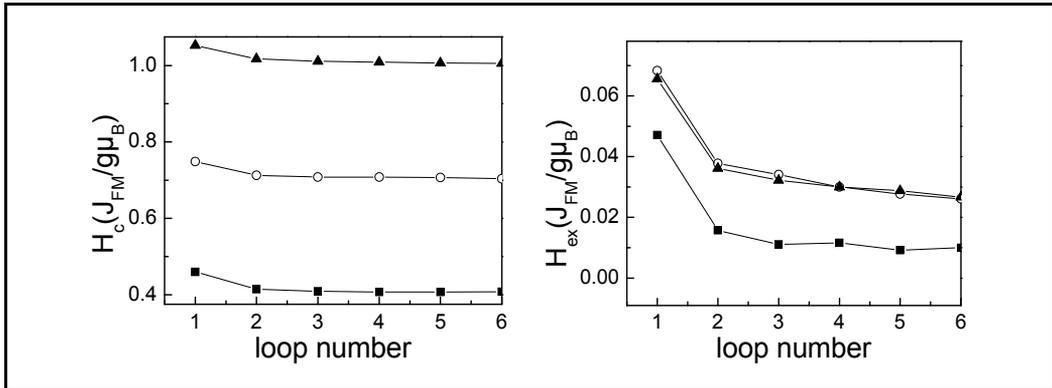


Fig. 9. Training effect of (a) $H_c$  and (b) $H_{\text{ex}}$  for core/shell nanoparticles with core size 5 lattice spacings and shell thickness 7(squares), 9(circles) and 15(triangles) respectively.

Nanogranular systems of which one of two phases, that are in contact, is spin-glass like phase are characterized by aging effects below their critical temperature ( $T_{\text{glass}}$ ), (Chamberlin et al., 1984) due to the slow response of the magnetisation with the time evolution. We have studied the aging effect, namely the slowing down of the spin dynamics with increasing the waiting time ( $t_w$ ) spent in the frozen state before any field variation for single nanoparticles with a FM core and a spin-glass like FI shell.

In figure 10 we present our MC simulation results for the time dependence of the thermoremanent magnetisation (TRM) for two different waiting times for a nanoparticle of radius 9 lattice spacings (Fiorani et al., 2006). The numerical procedure is the following: the system is cooled from a high temperature with zero field down to an initial temperature  $T_i = 0.75$  ( $J/k_B$ ), then a field  $H_{\text{cool}} = 0.4$  ( $J/g\mu_B$ ) is applied along the z-axis for different times  $t_w$  and we continue field cooling at a constant rate down to a low temperature  $T_f = 0.15$  ( $J_{\text{FM}}/k_B$ ). At this temperature the  $H_{\text{cool}}$  is removed and we calculate the TRM as a function of time expressed in MCS.

We observe that by increasing the waiting time, TRM decays slower with time because the system during the longer waiting time goes to a lower energy configuration and then by reducing the temperature the system remains trapped to this local minimum, as it is more difficult to overcome the energy barriers separating different states. Therefore the relaxation time of the TRM is enlarged.

The slowing down of the spin dynamics with increasing  $t_w$  has a noticeable effect on the exchange bias properties of single nanoparticles with FM core and a spin-glass like FI shell (Fiorani et al., 2006). Our MC studies demonstrated that the  $H_{\text{ex}}$  and  $M_r$  increase with  $t_w$ . This fact is important for technological applications, because by varying the waiting time we can control the exchange bias behaviour of the system.

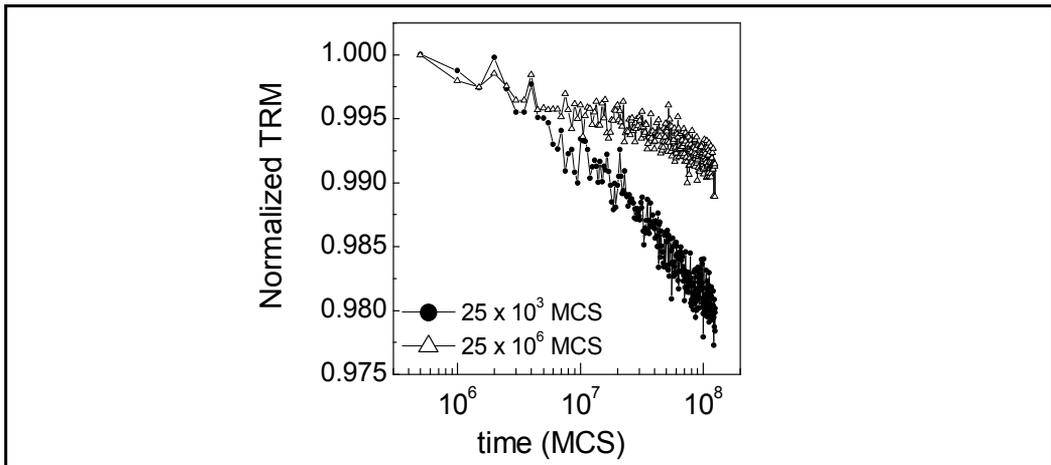


Fig. 10. TRM vs time for two different  $t_w$  values, as obtained by the MC simulations. TRM is normalized to the value at the beginning of the simulation.

## 6. MC simulation results of Interparticle interactions effects on magnetic nanoparticle assemblies

For decades assemblies of interacting magnetic nanoparticles have been a fascination and a challenge for materials scientists. Magnetic nanoparticles are commonly formed in assemblies, with either random or ordered structure. In the first group belong systems such as ferrofluids and granular solids, while in the second group belong the patterned media (or magnetic dots) and the self-assembled arrays of nanoparticles. In the assemblies of magnetic nanoparticles the crucial role of interparticle interactions in determining their response to an externally applied field as well as the temperature dependence of the magnetic properties has been recognized long ago (Dormann et al., 1997). In this article we review our results from MC simulations of the field and temperature dependence of the magnetisation of nanoparticle assemblies.

### 6.1 Random assemblies of magnetic nanoparticles

In a three dimensional (3D) random assembly of magnetic nanoparticles, especially at high densities, interparticle interactions have an important and sometimes dominant role in the formation of the magnetic behavior. Magnetostatic interactions between the particles are always present owing to the magnetic moment each particle carries. Due to their long range character they cannot be neglected except at the extreme dilute limit. Furthermore, exchange interaction between the particles appears when there is physical contact between them. The exchange interaction is expected to play an important role in samples with concentration close and above the percolation threshold. Indeed as the nanoparticle concentration increases, the interparticle interactions modify the distribution of the effective energy barrier, resulting in more complex phenomena, such as superspin glass (SSG) behaviour in low-enough temperatures for intermediate concentration systems (Sahoo et al., 2003) and superferromagnetic (SFM) order for very dense systems (Bedanta et al., 2007).

The characteristics of the hysteresis loop (remanence and coercivity) and the blocking temperature have been shown to vary with nanoparticle concentration in granular metals

and frozen ferrofluids (Dormann et al., 1997). The experimental trend was successfully reproduced by a model that includes interparticle dipolar interactions (Kechrakos & Trohidou, 1998; Chantrell et al., 2001).

We have investigated the role of interparticle exchange in modifying the concentration dependence of the hysteresis characteristics and the blocking temperature of a nanoparticle assembly (Kechrakos & Trohidou, 2003). The whole range of exchange constants strengths is studied, thus modelling the transition from well separated to coalesced particles.

The concentration dependence of the coercivity is shown in figure 11. In the dilute limit the coercivity at zero temperature is theoretically predicted to approach the value of  $H_c=0.96K_1/M_s$ . The data in figure 11 are slightly below this value due to the finite temperature ( $T=0.01 J/k_B$ ) at which the simulation is performed. Beyond the dilute limit, the effect of either type of interactions on the coercivity is quite similar, as the overall decrease of the coercivity values indicate. In the case of dipolar interaction only (full circles) the reduction of coercivity is due to the demagnetizing character of the dipolar forces that is also responsible for the reduction of the remanence. On the other hand, exchange interactions favor the formation of ferromagnetic clusters of particles with low anisotropy, as explained earlier, and therefore the magnetisation reversal is facilitated and the coercivity is reduced relative to the non-interacting case. When both types of interactions are present, we observe that their effect on the coercivity is not additive in all cases. In particular for weak exchange coupling (full circles and triangles) the interactions act cooperatively and the coercivity is further reduced relative to the case of exclusively dipolar or exchange forces. However, in the case of particle coalescence (open stars) the introduction of dipolar forces in the sample of coalesced particles (full stars) shifts the coercivity upwards. We attribute this behavior to the strong random dipolar fields generated in the sample containing large, almost isotropic, coherent clusters of particles. In this case, dipolar forces introduce an extra, albeit weak, anisotropy in the system that enhances the coercive field values.

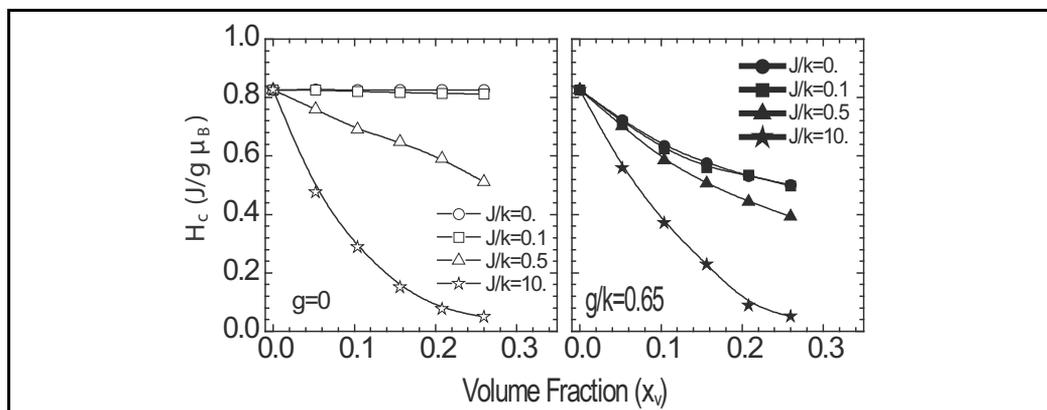


Fig. 11. Concentration dependence of coercivity at low temperature ( $T=0.01 J/k_B$ ), (a) exchange coupling only ( $g=0$ ), (b) exchange and dipolar coupling ( $g/k=0.65$ ).

We discuss next the temperature dependence of the magnetisation. The Zero-Field-Cooled / Field-Cooled (ZFC/FC) curves are calculated for metal volume fraction  $x_v = 0.15$ . Representative results are shown in figure 12. The peak of the ZFC curve provides the blocking temperature ( $T_b$ ) of the system (Dormann et al., 1997). Interparticle interactions produce an upshift of  $T_b$  and they modify the high temperature (superparamagnetic)

behavior of the magnetisation. In particular, Curie's law ( $\chi \sim 1/T$ ) that is valid in the case of non-interacting superparamagnetic particles (open circles), is no longer obeyed by the interacting system. The deviations from Curie's law are stronger when both types of interactions are present (full circles). In this case an almost linear decay with temperature is observed. Comparing the effect of the two types of interactions in the high temperature regime, we notice that dipolar interactions produce stronger deviations from Curie's law than exchange interactions and this is attributed to their long-range character. These results are in agreement with experimental findings on Fe nanoparticles in an Ag matrix (Binns et al., 2002).

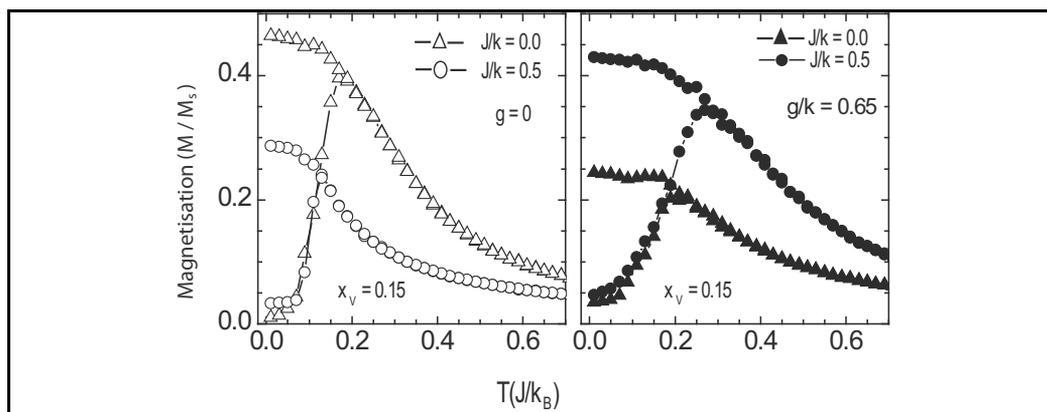


Fig. 12. Temperature dependence of magnetisation (ZFC/FC curves). (a) exchange coupling only ( $g=0$ ). (b) exchange and dipolar coupling ( $g/k=0.65$ ). Applied field  $h/k=0.1$ .

## 6.2 Ordered arrays of magnetic nanoparticles

Ordered arrays of magnetic nanoparticles (Petit et al., 1998; Murray et al., 2001; Punter et al., 2001) and patterned magnetic media (White, 2002) are currently the most promising materials for exploitation in high-density ( $\sim 1\text{Tb/in}^2$ ) magnetic storage media, due to the sharp distribution of their magnetic properties and their high reproducibility. Nanoparticle arrays (or super lattices) are prepared by colloidal synthesis followed by size-selective precipitation that produces a very narrow particle size distribution ( $\sigma < 5\%$ ). In addition to their technological applications, nanoparticles superlattices consist the ideal system for studying magnetisation reversal mechanism in the presence of interparticle interactions, due to the precise knowledge and control of the particle size and interparticle distances. Recent studies of self-assembled arrays of magnetic nanoparticles have provided clear evidence that interactions between the nanoparticles are present and manifest themselves in various aspects of their magnetic behaviour.

In figure 13 we show the zero-field cooled (ZFC) magnetisation curves for various coverage ( $c$ ) values and parameters corresponding to hard Co nanoparticles (Kechrakos & Trohidou, 2002). The maximum of the ZFC curve appears at the blocking temperature ( $T_b$ ) of the system (Dormann et al., 1997). An obvious increase of the blocking temperature with layer coverage is seen, that almost reaches saturation as soon as the first complete ML is formed ( $c=1$ ). The increase of  $T_b$  with coverage, below 1ML, is due to the anisotropic and ferromagnetic character of dipolar interactions that introduce an additional barrier to the magnetisation reversal of the particles. The increase of  $T_b$  with particle concentration has

also been observed in granular films (Dormann et al., 1997; Kechrakos & Trohidou, 1998; Chrantrell et al., 2001). Increased  $T_b$  values relative to the dilute limit have been recently measured in self-assembled Co nanoparticle arrays prepared from colloidal dispersions (Murray et al., 2001) and in self-organized lattices of Co clusters in  $Al_2O_3$  matrix (Luis et al., 2002). However, in the latter experiments the saturation of  $T_b$  values was found after 5-7 layers, while in closed-packed hexagonal arrays we demonstrate that saturation occurs already after two layers. Thus, we conclude that in hexagonal closed packed spherical Co particles the collective behaviour is predominantly determined by the intra-layer dipolar interactions, while inter-layer interactions play only a secondary role.

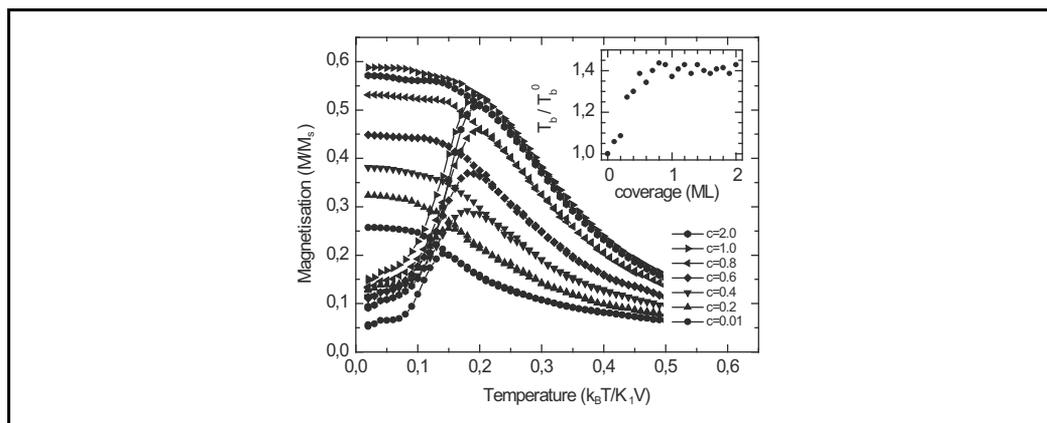


Fig. 13. Dependence of ZFC/FC magnetisation on layer coverage ( $c$ ) for hard Co nanoparticles ( $g/k=0.25$ ). Top layer is randomly occupied. In-plane applied field  $h/k=0.1$ . Inset shows the dependence of blocking temperature on monolayer coverage.

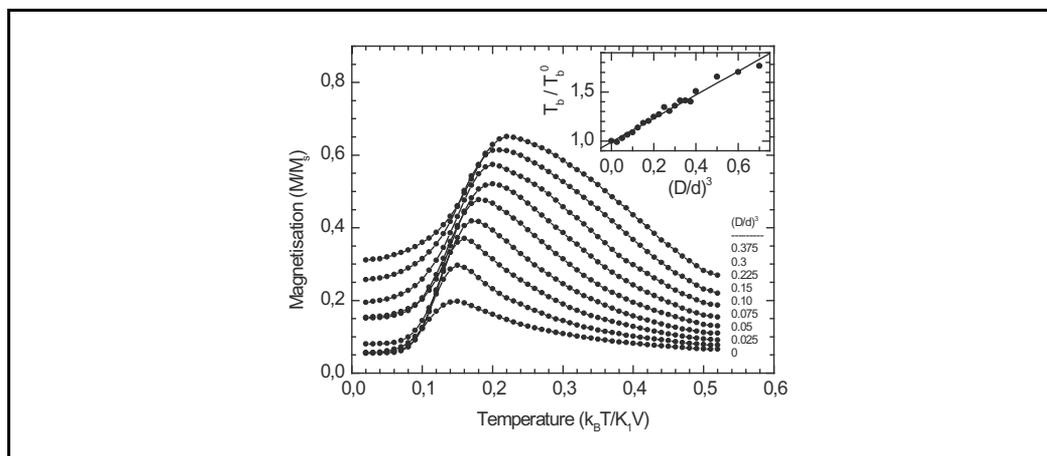


Fig. 14. Dependence of ZFC magnetisation on interparticle distance for hard Co nanoparticles ( $g/k=0.25$ ). Uppermost layer is randomly occupied. In-plane applied field  $h/k=0.1$ . Inset shows the scaling behaviour of blocking temperature with interparticle distance.

A strong dependence of the  $T_b$  on the interparticle distance ( $d$ ) is shown in figure 14. Our data indicate that for hard magnetic nanoparticles ( $g/k < 1$ ) the blocking temperature scales with the inverse cube of the interparticle distance,  $T_b \sim (D/d)^3$  (Kechrakos & Trohidou, 2002). Bearing in mind that the dipolar strength is  $g \sim (D/d)^3$ , we obtain  $T_b \sim g$ , which means that in hexagonal arrays of magnetically hard nanoparticles the dipolar interactions provide an additional energy barrier for magnetisation reversal which is proportional to the dipolar coupling strength. Our results are in qualitative agreement with recent measurements of ZFC/FC curves (Murray et al., 2001; Luis et al., 2002) indicating an increase of  $T_b$  relative to the dilute limit.

## 7. Concluding remarks

We have reviewed Monte Carlo simulation results on the magnetic behavior of nanoparticles and nanoparticle assemblies. Our results show that the MC simulation technique with the implementation of the Metropolis algorithm is an indispensable tool for the study of these systems, because they can include implicitly the details of the interface and the surface for the single nanoparticles and in all studied cases the temperature.

Our simulations have shown the dominant role of the surface to the magnetic behavior of antiferromagnetic and ferrimagnetic nanoparticles due to its anisotropy as well as the number of 'uncompensated' magnetic moments. The behavior of their coercive field  $H_c$  depends also on the details of their surface and on their size. The comparison between our simulation results and the experimental findings for antiferromagnetic nanoparticles NiO or the spherical and cubic ferrimagnetic nanoparticles  $\gamma$ -Fe<sub>2</sub>O<sub>3</sub> confirmed the important role of the surface and the uncompensated spins to the magnetic behavior of nanoparticles.

Our study was also extended to the case of composite nanoparticles with ferromagnetic core and a ferrimagnetic or antiferromagnetic shell. We demonstrated the important role of the interface in their magnetic behavior. The training and aging effects, which are characteristic of disordered systems, have been studied for the composite nanoparticles. The simulation results are in agreement with the experimental findings.

We demonstrated via Monte Carlo simulations in the case of the three-dimensional random assembly of interacting nanoparticles coupled via exchange and dipolar forces that the coercivity decreases with concentration for all values of the exchange strength and the blocking temperature increases with concentration, except when particle coalescence occurs and the system is above the percolation limit.

In the case of ordered arrays of nanoparticles our simulations showed that the blocking temperature increases with the coverage of the first monolayer and is rather insensitive to additional monolayers and decreases proportionally to the cube of the interparticle distance ( $T_b \sim 1/d^3$ ).

There are still several interesting issues to be addressed in the field of magnetic nanoparticles. The magnetic behaviour of complex inverted core/shell nanostructures. The influence of interparticle interactions on the magnetic properties of nanoparticles in dense nanoparticle assemblies is of major importance especially for the technological exploitation of magnetic nanoparticle assemblies. The interparticle interaction effects in nanoparticle assemblies with core/shell morphology are a new open field for study. MC simulations provide an extremely valuable tool for the study of these systems.

## 8. Acknowledgements

We would like to acknowledge the contribution of Dr. J.A. Blackman to parts of the work presented in this chapter and the helpful discussions with Dr. D. Fiorani, Prof. J. Nogués and Prof. C. Binns. This work was supported by the project of the Greek General Secretariat of Research PENED96, and the EC projects AMMARE (Contract No G5RD-CT-2001-00478) and NANOSPIN (Contract No NMP4-CT-2004-013545).

## 9. References

- Bader, S. D. (2006). Colloquium: Opportunities in nanomagnetism. *Reviews of Modern Physics*, Vol.78, No.1, (January 2006), pp. 1-15, ISSN 0034-6861
- Bahiana, M.; Pereira Nunes, J.P.; Altbir, D.; Vargas, P. & Knobel, M. (2004). Ordering effects of the dipolar interaction in lattices of small magnetic particles. *Journal of Magnetism and Magnetic Materials*, Vol.281, No.2-3, (October 2004), pp. 372-377, ISSN 0304-8853
- Baker, C.; Hasanain, S. K. & Shah, S. I. (2004). The magnetic behavior of iron oxide passivated iron nanoparticles. *Journal of Applied Physics*, Vol.96, No.11, (December 2004), pp. 6657-6666, ISSN 0021-8979
- Bedanta, S.; Eimuller, T.; Kleemann, W.; Rhensius, J.; Stromberg, F.; Amaladass, E.; Cardoso, S.; Freitas, P.P. (2007). Overcoming the dipolar disorder in dense CoFe nanoparticle ensembles: Superferromagnetism. *Physical Review Letters*, Vol.98, No. 17, (April 2007), pp. 176601, ISSN 1079-7114
- Bhowmik, R. N.; Nagarajan, R. & Ranganathan, R. (2004). Magnetic enhancement in antiferromagnetic nanoparticle of CoRh<sub>2</sub>O<sub>4</sub>. *Physical Review B*, Vol.69, No.5, (February 2004), pp. 054430-5, ISSN 1098-0121
- Binder, K.; Rauch, H. & Wildpaner, H. (1970). Monte Carlo calculation of the magnetisation of superparamagnetic particles. *Journal of Physics and Chemistry of Solids*, Vol.31, No.2, (February 1970), pp. 391-397, ISSN 0022-3697
- Binder, K. (1986). *Topics in Current Physics, Monte Carlo Methods in Statistical Physics*, Springer-Verlag, ISBN: 3-540-16514-2, Berlin Heidelberg New York Tokyo
- Binder, K. & Young, A. P. (1986). Spin glasses: Experimental facts, theoretical concepts, and open questions. *Reviews of Modern Physics*, Vol.58, No.4, (October-December 1986), pp. 801-976, ISSN 0034-6861
- Binder, K. (1987). *Applications of the Monte-Carlo Method in Statistical Physics*, Springer-Verlag, ISBN: 3540176500, New York
- Binns, C.; Maher, M.J.; Pankhurst, Q. A.; Kechrakos, D. & Trohidou, K.N. (2002). Magnetic behavior of nanostructured films assembled from preformed Fe clusters embedded in Ag. *Physical Review B*, Vol.66, No.18, (November 2002), pp. 184413-12, ISSN 1098-0121
- Bødker, F.; Mørup, S. & Linderøth, S. (1994). Surface effects in metallic iron nanoparticles. (1994) Surface effects in metallic iron nanoparticles. *Physical Review Letters*, Vol.72, No.2, (January 1994), pp. 282-285, ISSN 1079-7114
- Brown, G.; Janotti, A.; Eisanbach, M. & Stocks, G. M. (2005). Intrinsic volume scaling of thermally induced magnetisation in antiferromagnetic nanoparticles. *Physical Review B*, Vol.72, No.14, (October 2005), pp. 140405R-4, ISSN 1098-0121

- Chamberlin, R.V.; Mozurkewich, V.; Orbach, R. (1984). Time Decay of the Remanent Magnetisation in Spin-Glasses. *Physical Review Letters*, Vol.52 No.10, (March 1984) , pp. 867-870, ISSN 1079-7114
- Chen, C.; Kitakami, O. & Shimada, Y. (1998). Particle size effects and surface anisotropy in Fe-based granular films. *Journal of Applied Physics*, Vol.84, No.4, (August 1998), pp. 2184-2188, ISSN 0021-8979
- Chantrell, R. W.; Walmsley, N., Gore, J. & Maylin, M. (2001). Calculations of the susceptibility of interacting superparamagnetic particles. *Physical Review B*, Vol.63, No.2, (January 2001), pp. 024410-14, ISSN 1098-0121
- Coey, J. M. D. (1971). Noncollinear Spin Arrangement in Ultrafine Ferrimagnetic Crystallites. *Physical Review Letters*, Vol.27, No.17, (October 1971), pp. 1140-1142, ISSN 1079-7114
- Curiale, J.; Granada, M. ; Troiani, H. E. ; Sánchez, R. D. ; Leyva, A. G. ; Levy, P. & Samwer, K. (2009). Magnetic dead layer in ferromagnetic manganite nanoparticles. *Applied Physics Letters*, Vol.95, No.4, (July 2009), pp. 043106-3, ISSN 0003-6951
- De Biasi, E.; Ramos, C. A. ; Zysler, R. D. & Romero, H. (2002). Large surface magnetic contribution in amorphous ferromagnetic nanoparticles. *Physical Review B*, Vol.65, No.14, (March 2002), pp. 144416-8, ISSN 1098-0121
- Del Bianco, L.; Fiorani, D. ; Testa, A. M.; Bonetti, E. ; Savini, L. & Signoretti, S. (2003). Magnetic properties of the Fe/Fe oxide granular system. *Journal of Magnetism and Magnetic Materials*, Vol.262, No.1, (May 2003), pp. 128-131, ISSN 0304-8853
- Del Bianco, L. D.; Fiorani, D.; Testa, A. M.; Bonetti, E. & Signorini, L. (2004). Field-cooling dependence of exchange bias in a granular system of Fe nanoparticles embedded in an Fe oxide matrix. *Physical Review B*, (August 2004), pp. 052401 -4, ISSN 1098-0121
- Dimitrov, D. A. & Wysin, G. M. (1994). Effects of surface anisotropy on hysteresis in fine magnetic particles. *Physical Review B*, Vol.50, No.5, (August 1994), pp. 3077-3084, ISSN 1098-0121
- Dimitrov, D. A. & Wysin, G. M. (1996). Magnetic properties of superparamagnetic particles by a Monte Carlo method. *Physical Review B*, Vol.54, No.13, (October 1996), pp. 9237-9241, ISSN 1098-0121
- Dormann, J. L., Fiorani, D. and Tronc, E. (1997). *Magnetic Relaxation in Fine-Particle Systems, in Advances in Chemical Physics*, Vol.98, pp. 283-494 (eds I. Prigogine and S. A. Rice), John Wiley & Sons, Inc., Hoboken, New York, USA Book Series: Advances in Chemical Physics, ISBN: 9780471162858
- Du, H. F. & Du, A. (2006). The hysteresis curves of nanoparticles obtained by Monte Carlo method based on the Stoner-Wohlfarth model. *Journal of Applied Physics*, Vol.99, No.10, (May 2006), pp. 104306-4, ISSN 0021-8979
- Eftaxias, E. & Trohidou, K. N. (2005). Numerical study of the exchange bias effects in magnetic nanoparticles with core / shell morphology. *Physical Review B*, Vol.71, No.13, (April 2005), pp. 134406-6, ISSN 1098-0121
- Eftaxias, E.; Vasilakaki, M. & Trohidou, K. N. (2007). A Monte Carlo study of the exchange bias effects in magnetic nanoparticles with ferromagnetic core/ antiferromagnetic shell morphology. *Modern Physics Letters B*, Vol.21, No. 21, (August 2007), pp. 1169-1177, ISSN: 0217-9849
- Fiorani, D., Del Bianco, L.; Testa, A.M. & Trohidou, K.N. (2006). Glassy dynamics in the exchange bias properties of the iron/iron oxide nanogranular system. *Physical Review B*, Vol.73, No.9, (March 2006), pp. 092403, ISSN 1098-0121

- Gangopadhyay, S.; Hadjipanayis, G. C. ; Dale, B. ; Sorensen, C. M. ; Klabunde, K. J. ; Papaefthymiou, V. & Kostikas, A. (1992). Magnetic properties of ultrafine iron particles. *Physical Review B*, Vol.45, No.17, (May 1992), pp. 9778-9787, ISSN 1098-0121
- Garanin, D. A. & Kachkachi, H. (2003). Surface Contribution to the Anisotropy of Magnetic Nanoparticles. *Physical Review Letters*, Vol.90, No.6, (February 2003), pp. 065504-4, ISSN 1079-7114
- Garcia-Otero, J.; Porto, M.; Rivas, J. & Bunde, A. (1999). Influence of the cubic anisotropy constants on the hysteresis loops of single-domain particles: A Monte Carlo study. *Journal of Applied Physics*, Vol.85, No.4, (February 1999), pp. 2287-2292, ISSN 0021-8979
- Grzybowski, A.; Gwóźdź, E. & Bródka, A. (2000). Ewald summation of electrostatic interactions in molecular dynamics of a three-dimensional system with periodicity in two directions. *Physical Review B*, Vol.61, No.10, (March 2000), pp. 6706-6712, ISSN 1098-0121
- Hinzke, D. & Nowak., U. (1999). Monte Carlo simulation of magnetisation switching in a Heisenberg model for small ferromagnetic particles. *Computer Physics Communications*, Vol.121-122, (September-October 1999), pp. 334-337, ISSN 0010-4655
- Iglesias, O. & Labarta, A. (2004). Shape and surface anisotropy effects on the hysteresis of ferrimagnetic nanoparticles. *Journal of Magnetism and Magnetic Materials*, Vol.272-276, No.1, (May 2004), pp. 685-686, ISSN 0304-8853
- Jamet, M.; Wernsdorfer, W.; Thirion, C.; Maily, D. ; Dupuis, V. ; Mélinon, P. & Pérez, A. (2001). Magnetic Anisotropy of a Single Cobalt Nanocluster. *Physical Review Letters*, Vol.86, No.20, (June 2001), pp. 4676-4679, ISSN 1079-7114
- Kachkachi, H.; Nogues, M.; Tronc, E. & Garanin, D. A. (2000). Finite-size versus surface effects in nanoparticles. *Journal of Magnetism and Magnetic Materials*, Vol.221, No.1-2, (November 2000), pp. 158-163, ISSN 0304-8853
- Kechrakos, D. & Trohidou, K. N. (1998). Magnetic properties of dipolar interacting single-domain particles. *Physical Review B*, Vol.58, No.18, (November 1998), pp. 12169-12177, ISSN 1098-0121
- Kechrakos, D. & Trohidou, K. N. (2002). Magnetic properties of self-assembled interacting nanoparticles. *Applied Physics Letters*, Vol.81, No.24, (December 2002), pp. 4574-4576, ISSN 0003-6951
- Kechrakos, D. & Trohidou, K. N. (2003). Competition between dipolar and exchange interparticle interactions in magnetic nanoparticle films. *Journal of Magnetism and Magnetic Materials*, Vol.262, No.1, (May 2003), pp. 107-110, ISSN 0304-8853
- Kneller, E. F. & Luborsky, F. E. (1963). Particle Size Dependence of Coercivity and Remanence of Single-Domain Particles. *Journal of Applied Physics*, Vol.34, No.3, (March 1963), pp. 656-659, ISSN 0021-8979
- Kodama, R. H. (1999). Magnetic nanoparticles. *Journal of Magnetism and Magnetic Materials*, Vol.200, No.1-3, (October 1999), pp. 359-372, ISSN 0304-8853
- Kretschmer, R. & Binder, K. (1979). Ordering and phase transitions in Ising systems with competing short range and dipolar interactions. *Zeitschrift für Physik B*, Vol.34, No.4, (December 1979), pp. 375-392, ISSN 0722-3277

- Landau, D. P. (1976). Finite-size behavior of the Ising square lattice. *Physical Review B*, Vol.13, No.7, (April 1976), pp. 2997-3011, ISSN 1098-0121
- Landau, D. P. & Binder, K. (2000). *A Guide to Monte Carlo Simulations in Statistical Physics*, Cambridge University Press, ISBN: 0 521 65314 2, U. K.
- Leighton, C.; Nogues, J. ; Jonsson-Akerman, B. J. & Schuller, I. K. (2000). Coercivity Enhancement in Exchange Biased Systems Driven by Interfacial Magnetic Frustration. *Physical Review Letters*, Vol.84, No.15, ( April 2000), pp. 3466-3469, ISSN 1079-7114
- Leite, V. S.; Godoy, M. & Figueiredo, W. (2005). Finite-size effects and compensation temperature of a ferrimagnetic small particle. *Physical Review B*, Vol.71, No.9, (March 2005), pp. 094427 -7, ISSN 1098-0121
- Lin, X.; Murthy, A. S.; Hadjipanayis, G. C. ; Swann, C. & Shah, S. I. (1994). Magnetic and structural properties of Fe-FeO bilayers. *Journal of Applied Physics*, Vol.76, No.10, (November 1994 ), pp. 6543 -6546, ISSN 0021-8979
- Luis, F.; Petroff, F.; Torres, J.M.; Garcia, L.M.; Bartolome, J.; Carrey, J. & Vaures, A.(2002). Magnetic relaxation of interacting Co clusters: Crossover from two- to three-dimensional lattices. *Physical Review Letters*, Vol.88, No.21, (May 2002), pp. 217205, ISSN 0031-9007
- Martinez, B.; Obradors, X. ; Balcells, Ll.; Rouanet, A. & Monty, C. (1998). Low Temperature Surface Spin-Glass Transition in  $\gamma$ -Fe<sub>2</sub>O<sub>3</sub> Nanoparticles. *Physical Review Letters*, Vol. 80, No.1, (January 1998), pp. 181-184, ISSN 1079-7114
- Meiklejohn, W. H. & Bean, C. P. (1957). New Magnetic Anisotropy. *Physical Review*, Vol. 105, No.3, (February 1957), pp. 904-913, ISSN 0031-899X
- Metropolis, N.; Rosenbluth, A.; Rosenbluth, M.; Teller, A. & Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, Vol.21, No.6, (June 1953), pp. 1087-1092, ISSN 0021-9606
- Morel, R.; Brenac, A. & Portemont, C. (2004). Exchange bias and coercivity in oxygen-exposed cobalt clusters. *Journal of Applied Physics*, Vol.95, No.7, (April 2004), pp. 3757-3761, ISSN 0021-8979
- Mørup, S. & Hansen, B. R. (2005). Uniform magnetic excitations in nanoparticles. *Physical Review B*, Vol.72, No.2, (July 2005), pp. 024418-6, ISSN 1098-0121
- Mørup, S.; Madsen, D. E.; Frandsen, C.; Bahl C. R. H. & Hansen, M. F. (2007). Experimental and theoretical studies of nanoparticles of antiferromagnetic materials. *Journal of Physics: Condensed Matter*, Vol.19, No.21, (May 2007), pp. 213202 -212333, ISSN 0953-8984
- Murray, C. B.; Sun, S.H.; Doyle, H.; Betley, T. (2001). Monodisperse 3d transition-metal (Co, Ni, Fe) nanoparticles and their assembly into nanoparticle superlattices. *Materials Research Society Bulletin*, Vol.26, No.12, (December 2001), pp. 985-991
- Néel, L. (1953). Some New Results on Antiferromagnetism and Ferromagnetism. *Reviews of Modern Physics*, Vol.25, No.1, (January-March 1953), pp. 58-63, ISSN 1539-0756
- Nogués, J. & Schuller, I. K. (1999). Exchange bias. *Journal of Magnetism and Magnetic Materials*, Vol.192, No.2, (February 1999), pp. 203-232, ISSN 0304-8853
- Nogués, J.; Sort, J.; Langlais, V.; Skymryev, V.; Surinach, S.; Munoz, J. & Baro, M. D. (2005). Exchange bias in nanostructures. *Physics Reports*, Vol.422, No.3, (December 2005), pp. 65-117, ISSN: 0370-1573

- Nogués, J.; Skumryev, V.; Sort, J.; Stoyanov, S. & Givord, D. (2006). Shell-Driven Magnetic Stability in Core-Shell Nanoparticles. *Physical Review Letters*, Vol.97, No.15, (October 2006), pp. 157203-4, ISSN 1079-7114
- Pankhurst, Q. A.; Connolly, J.; Jones, S. K. & Dobson, J. (2003). Applications of magnetic nanoparticles in biomedicine. *Journal of Physics D: Applied Physics*, Vol.36, No.13, (June 2003), pp. R167-181, ISSN 0022-3727
- Passamani, E. C.; Larica, C.; Marques, C.; Proveti, J. R.; Takeuchi, A. Y. & Sanchez, F. H. (2006). Exchange bias and anomalous vertical shift of the hysteresis loops in milled Fe/MnO<sub>2</sub> material. *Journal of Magnetism and Magnetic Materials*, Vol.299, No.1, (April 2006), pp. 11-20, ISSN 0304-8853
- Peng, D.L.; Hihara, T.; Sumiyama, K.; Morikawa, H. (2002). Structural and magnetic characteristics of monodispersed Fe and oxide-coated Fe cluster assemblies. *Journal of Applied Physics*, Vol.92, No.6, (September 2002), pp. 3075-3083, ISSN 0021-8979
- Petit, C.; Taleb, A.; Pileni, M.P. (1998). Self-organization of magnetic nanosized cobalt particles. *Advanced Materials*, Vol.10, No.3, (February 1998), pp. 259-261, ISSN 1521-4095
- Puntes, V. F.; Krishnam, K. M. & Alivisatos, A. P. (2001). Synthesis, self-assembly, and magnetic behavior of a two-dimensional superlattice of single-crystal  $\epsilon$ -Co nanoparticles. *Applied Physics Letters*, Vol.78, No.15, (April 2001), pp. 2187-2190, ISSN 0003-6951
- Respaud, M.; Broto, J. M.; Rakoto, H.; Fert, A. R.; Thomas, L.; Barbara, B.; Verelst, M.; Snoeck, E.; Lecante, P.; Mosset, A.; Osuna, J.; Ould Ely, T.; Amiens, C. & Chaudret, B. (1998). Surface effects on the magnetic properties of ultrafine cobalt particles. *Physical Review B*, Vol.57, No.5, (February 1998), pp. 2925-2935, ISSN 1098-0121
- Sahoo, S.; Petravic, O.; Kleemann, W.; Stappert, S.; Dumpich, G.; Nordblad, P.; Cardoso, S.; Freitas, P.P. (2003). Cooperative versus superparamagnetic behavior of dense magnetic nanoparticles in Co<sub>80</sub>Fe<sub>20</sub>/Al<sub>2</sub>O<sub>3</sub> multilayers. *Applied Physics Letters*, Vol. 82, No.23, (June 2003), pp. 4116-4118, ISSN 0003-6951
- Sinohora, T.; Sato, T.; Taniyama, T. & Nakatani, I. (1999). Size dependent magnetisation of PdFe fine particles. *Journal of Magnetism and Magnetic Materials*, Vol.196-197, (May 1999), pp. 94-95, ISSN 0304-8853
- Skumryev, V.; Stoyanov, S.; Zhang, Y.; Hadjipanayis, G.; Givord, D. & Nogués J. (2003). Beating the superparamagnetic limit with exchange-bias. *Nature (London)*, Vol.423, No.6942, (June 2003), pp. 850-853, ISSN 0028-0836
- Stoner, E. C. & Wohlfarth, E. P. (1948). A Mechanism of Magnetic Hysteresis in Heterogeneous Alloys. *Philosophical Transactions of the Royal Society London, Series A*, Vol.240, No.826, (May 1948), pp. 599 - 642, Online ISSN: 1471-2962
- Trohidou, K. N.; Zianni, X. & Blackmann, J. A. (1998). Surface effects on the magnetic behavior of antiferromagnetic particles. *Journal of Applied Physics*, Vol.84, No.5, (September 1998), pp. 2795-2801, ISSN 0021-8979
- Trohidou, K. N. (2005). Monte Carlo studies of surface and interface effects in magnetic nanoparticles, In: *Surface Effects in Magnetic Nanoparticles (Nanostructured science and technology)*, Fiorani, D., pp. 45-74, Springer, ISBN 0-387-23279-6, N. York, U. S. A.
- Trohidou, K.N.; Vasilakaki, M.; Del Bianco, L.; Fiorani, D. & Testa, A.M. (2007). Exchange bias in a magnetic ordered/disordered nanoparticle system A Monte Carlo

- simulation study. *Journal of Magnetism and Magnetic Materials*, Vol.316, No.2, (September 2007), pp. e82-e87, ISSN 0304-8853
- Vasilakaki, M. & Trohidou, K. N. (2008). Surface effects on the magnetic behavior of nanoparticles with core / shell morphology. *Journal of Physics D: Applied Physics*, Vol.41, No.13, (July 2008), pp. 134006-5, ISSN 0022-3727
- Vasilakaki, M. & Trohidou, K. N. (2009). Numerical study of the exchange-bias effect in nanoparticles with ferromagnetic core / ferrimagnetic disordered shell morphology. *Physical Review B*, Vol.79, No.14, (April 2009), pp. 144402-8, ISSN 1098-0121
- Wildpaner, V. (1974). Monte Carlo study of small magnetic particles. *Zeitschrift für Physik A*, Vol.270, No.3, (February 1974), pp. 215-223, ISSN 0939-7922
- Winkler, E.; Zysler, R. D.; Vasquez Mansilla, M.; Fiorani, D.; Rinaldi, D.; Vasilakaki, M. & Trohidou, K. N. (2008). Surface spin-glass freezing in interacting core-shell NiO nanoparticles. *Nanotechnology*, Vol.19, No.18, (May 2008), pp. 185702-185710, ISSN 0957-4484
- White, R.L. (2002). Magnetisation processes in patterned media. *Journal of Magnetism and Magnetic Materials*, Vol.242, No.1, (April 2002), pp. 21-26, ISSN 0304-8853
- Zianni, X. & Trohidou, K. N. (1998). Monte Carlo simulations on the coercive behaviour of oxide coated ferromagnetic particles. *Journal of Physics: Condensed Matter*, Vol.10, No.33, (August 1998), pp. 7475-7483, ISSN 0953-8984

# Monte Carlo Simulation for Magnetic Domain Structure and Hysteresis Properties

Katsuhiko Yamaguchi, Kenji Suzuki and Osamu Nittono  
*Fukushima University*  
*Japan*

## 1. Introduction

Recently many studies for magnetic process simulations of micro magnetic clusters have been performed using several calculation methods. These studies are expected to be available to realize high-density magnetic memories, new micro-magnetic devices or to analyze microscopically for magnetic non destructive evaluation. Monte Carlo (MC) method is one of useful and powerful methods to simulate magnetic process for magnetic clusters including complicated interaction such as different exchange interactions due to different elements and to introduce magnetic properties depending on temperature.

To apply MC method for magnetic process simulation, there were some problems. One is that MC method is originally dealing with stable states, that is, the time processes on MC simulations can not be usually recognized as the real changes on time, e.g. for hysteresis curves (M-H curves) with increasing and decreasing applied magnetic field. Then a pseudo-dynamic process for MC method is introduced for dealing with such a simulation on section 2. Next problem is that the MC calculation for large clusters demands huge CPU time because it is necessary to repeat MC step (MCS) until  $N$  for the cluster cell number  $N$ . Especially the magnetic dipole interaction which is included in Hamiltonian must be calculated among all the spins in the cluster. Then a new technique of MC method by a parallelized program is introduced for dealing with larger cluster on section 3. The useful calculation results using these MC methods are presented on following sections. Section 4 introduces the producing of magnetic domains and domain walls (DWs) for the clusters including spins affected by exchange interaction, magnetic dipole interaction and crystal anisotropy. On section 5, magnetic domain wall displacements (DWDs) are shown for nano-wires with local magnetic impurity. On section 6, M-H curves are shown for magnetic clusters with a local magnetic distribution corresponding with grain boundary of Ni based alloy. For elementary theory on MC method, previous chapter should be referred.

## 2. Pseudo-dynamic process on MC method

In general, MC method deals with thermal equilibrium states. Therefore usually MC steps are repeated until getting a stable state. Here 1 MC step (MCS) means scanning up to the total cell number of times for the spin-flip process. Ordinary repeating MCS is set to  $N$  MCS,

here  $N$  is the total number of spin sites. But now we stopped the repeating before getting a stable state because of dealing with magnetic dynamic processes (Yamaguchi *et al.* 2004). Under the constant magnetic field condition, the total spin is in a non-equilibrium state and going to an equilibrium state with progressing MC steps. The magnetic field slightly increases before achievement of the equilibrium state, then the total spin is kept under another non-equilibrium state again and proceeding to a new equilibrium state as show Fig.1. The operation is renewed until achievement of final magnetic field. Because the change of the magnetic field is minute, it will be able to regard approximately that a series of steps is continuous process through a pseudo-non-equilibrium state. Here an assumption is introduced that magnetization intensity, namely the summation of total spin, of each MC step can reflect the magnetic dynamic process on magnetic hysteresis.

Pseudo-dynamic process on MC method is useful for dealing with magnetic dynamic simulation, e.g., magnetic hysteresis curves or magnetic domain wall moving, as they are explained in later sections.

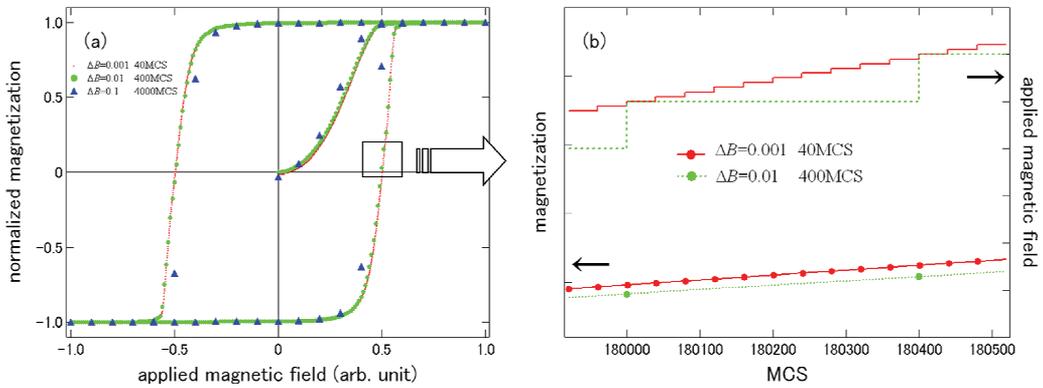


Fig. 1. (a) Magnetic hysteresis curves for a cluster with different step of applied magnetic field  $\Delta B$ . (b) Example of MC step dependence on applied magnetic field and magnetization. Circles show the last data of magnetization under the same condition.

### 3. Parallelized MC algorithm

In this section, for explanation of parallelized MC algorithm, a following Hamiltonian is used:

$$H = H_J + H_D + H_B$$

$$= - \sum_{near} J_{ij} \mathbf{S}_i \cdot \mathbf{S}_j + D \sum_{all} \left( \frac{\mathbf{S}_i \cdot \mathbf{S}_j}{|\mathbf{r}_{ij}|^3} - \frac{3}{|\mathbf{r}_{ij}|^5} (\mathbf{S}_i \cdot \mathbf{r}_{ij})(\mathbf{S}_j \cdot \mathbf{r}_{ij}) \right) + B \sum_i \mathbf{S}_i. \quad (1)$$

$H_J$  term,  $H_D$  term and  $H_B$  term represent exchange interaction energy, magnetic dipole interaction energy and applied magnetic field energy, respectively. Here  $\mathbf{S}_i$  denotes the spin state of  $i$ -th cell and  $\mathbf{r}_{ij}$  represents the distance between  $i$ -th spin and  $j$ -th spin. Below we deal with clusters with the lattice constant of 1 and this is regarded as a criterion of length. In the first term  $H_J$ ,  $J_{ij}$  stands for an exchange interaction energy constant for  $i$ -th and  $j$ -th spins.

Usually exchange interaction works on only neighbor spins, because the interaction is originally due to overlapping between wave functions of electrons with spins, then the summation is limited to the extent in an effective radius  $r_{eff}$  from a target spin  $S_i$ :  $|r_{ij}| \leq r_{eff}$ . In the second term  $H_D$ ,  $D$  for a magnetic dipole interaction constant for  $i$ -th and  $j$ -th spins. The magnetic dipole interaction works on all spins because it is due to magnetic field interspersed in all space. Then the summation includes the interaction energy between  $i$ -th spin and all  $j$ -th spins except for  $j=i$ . In the third term  $H_B$ ,  $B$  represents applied magnetic field which acts equally all spins.

For parallelizing MC program, it is important to keep causality of MC algorithm. Hence it is not allowed that before a spin  $S_i$  is updated by MC process, the next calculation starts about another spin  $S_i'$ . Therefore a feasible parallelized process is limited to the summation for a fixed  $S_i$ . Then Eq.(1) was transformed for applying the parallelized algorithm to MC method without spoiling the causality as follows:

$$H = \sum_i \left\{ \sum_{j \neq i} \left[ -J_{ij} \mathbf{S}_i \cdot \mathbf{S}_j \delta_{|i-j|,1} + D \left( \frac{\mathbf{S}_i \cdot \mathbf{S}_j}{|\mathbf{r}_{ij}|^3} - \frac{3}{|\mathbf{r}_{ij}|^5} (\mathbf{S}_i \cdot \mathbf{r}_j)(\mathbf{S}_j \cdot \mathbf{r}_i) \right) \right] + B \mathbf{S}_i \right\}. \quad (2)$$

Here the inner summation for  $j$  of Eq.(2) can be parallelized. Kronecker's  $\delta$  limits the summation of  $j$  for the first term to the extent of the nearest neighbors (note  $r_{eff}=1$  in this case) with checking the distance between  $i$ -th and  $j$ -th spins on each selection of a target spin  $S_i$ . Although the check process adds a load for CPU power, the program parallelizing the summation of  $j$  in block is effective for larger clusters.

Figure 2 shows a flowchart of the MC algorithm including the parallelized process. After choosing a target spin  $S_i$  randomly under an initial state, all  $j$ -th spins except for  $j=i$  are divided into plural CPU in a parallel computer. A CPU assigned to a set for  $S_i$  and  $S_j$  calculates  $r_{ij}$  and distinguishes  $|r_{ij}| \leq r_{eff}$  and  $|r_{ij}| > r_{eff}$ . Note that  $r_{eff} \geq 1$  is allowed in general. The CPU calculates  $H_j$  and  $H_D$ , and the summation of them is stocked into a memory with the results by other CPUs. This process is repeated until last  $j$  ( $=N$ ) which is the total spin number of dealing cluster. After adding applied magnetic field energy  $H_B$ , the target spin  $S_i$  is updated by Metropolis method (Metropolis *et al.* 1953, Landau & Binder 2000). The update of  $S_i$  is repeating  $N$  times, that is, all spins are updated as an average. This period is called one MC step (1 MCS). For getting stable physical quantities, the calculation process is repeating  $M$  times ( $= M$  MCS) under the same condition.  $M$  sets usually  $N$ , therefore the parallelized process repeats  $N^2$  times and the process is expected to reduce the calculation time. Using above algorithm, all simulations in this chapter were carried out by the use of the parallel super-computer, Altix3700B in the Institute of Fluid Science, Tohoku University (Japan).

Figure 3 shows the wall time (actual calculating time) during 1000 MCS repeating for different size squares with the one side length  $L=20, 30, 50, 75, 100$  and  $150$  cells for each CPU number used in the same time.  $N$  ( $=L^2$ ) is total cell number. The increase of CPU number effectively reduces the calculation time especially for larger clusters. The calculation results for the same cluster have no discrepancy among using of different CPU numbers.

Figure 4 shows the total CPU time and the wall time for the calculations for different size clusters at a fixed temperature. The numbers in brackets show the CPU numbers for each calculation.

Figure 5 shows results of temperature dependence of the normalized magnetization  $M$  for different size clusters. For clusters with the one side length between  $L=10$  and  $50$ , the results well obey the Curie-Weiss law and the Curie temperatures were estimated at about  $k_B T_c=1.0$ . For larger clusters, however, the increases of the magnetizations are not seen at low temperature.

In general it is known that closure domain structure of spin system appears for thin film magnetic cluster due to magnetic dipole interaction although single magnetic domain is produced for the smaller cluster (Sasaki & Matsubara 1997, Vedmedenko *et al.* 2000). Then above results of magnetization will be also size effect due to magnetic dipole interaction.

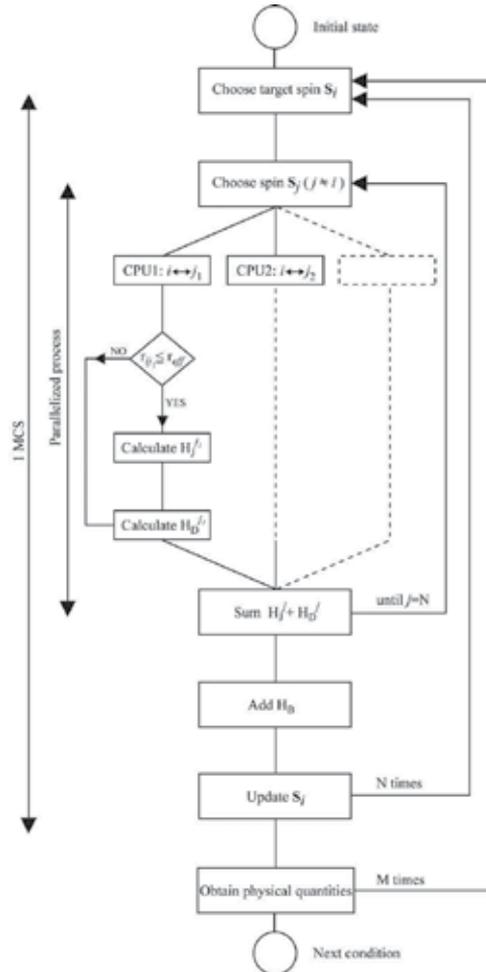


Fig. 2. Flowchart of MC algorithm including parallelized process. The process from “Choose spin  $S_j^i$ ” to “Sum  $H_j + H_D$ ” is parallelized in this algorithm. The process from “Choose spin  $S_i^j$ ” to “Update  $S_i^j$ ” is repeating until spin total number  $N$  and it is called 1MCS.

Figure 6 shows spin snapshots for the different size square clusters with the one side length of  $L=10, 50, 75$ , respectively at lowest temperature. It is clearly seen that the closure domain structure of spin system actually appears for the cluster with  $L=75$ .

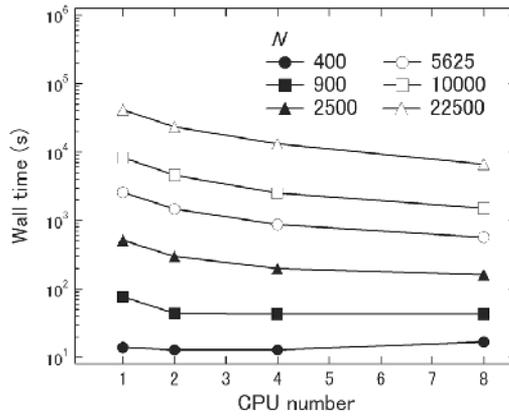


Fig. 3. Wall time during 1000 MCS depending on CPU number for each size cluster ( $N=L^2$ ).

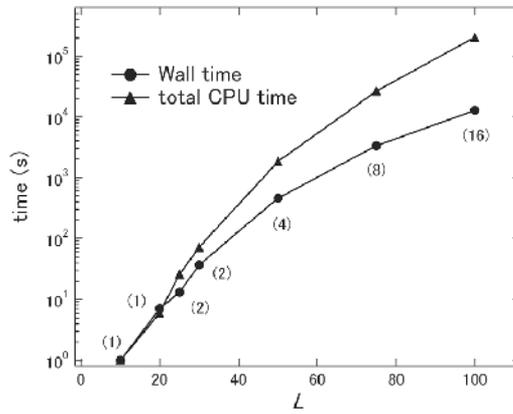


Fig. 4. Total CPU time and wall time on calculation at a fixed temperature for each size cluster. Numbers in brackets ( ) show the CPU numbers for parallel calculation.

The closure domain structure parameters  $M_\phi$  for different size square clusters are shown in Fig.6. Here  $M_\phi$  is given by equation as below,

$$M_\phi = \frac{1}{N} \sum_i \left( \mathbf{S}_i \times \frac{\mathbf{r}_i - \mathbf{r}_c}{|\mathbf{r}_i - \mathbf{r}_c|} \right)_z \tag{3}$$

$N$  represents total spin number and  $r_i$  and  $r_c$  are coordinate vectors of the spin  $S_i$  and the center of circle structure, respectively. Figure 6 shows  $M_\phi$  increases as temperature decreases for the cluster with  $L=75$  and  $100$ .

Figure 7 shows the variation of normalized magnetization  $M$  and the closure domain structure parameter  $M_\phi$  depending on size of square clusters with the one side length  $L$ . It is clearly seen that single domain structure turns to the closure domain structure accompanied with increasing of  $L$ .

As a result, the parallelized algorithm is available for the greater clusters including magnetic dipole interaction.

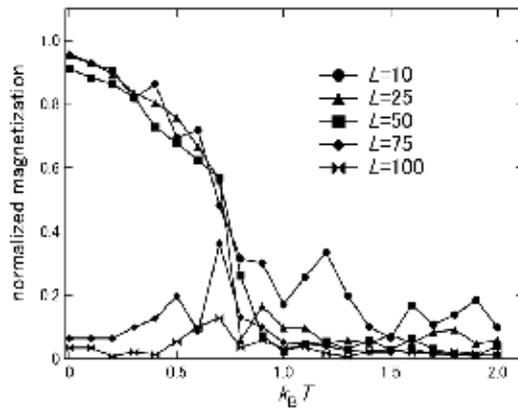


Fig. 5. Temperature dependence of normalized magnetization  $M$  for different size square clusters.

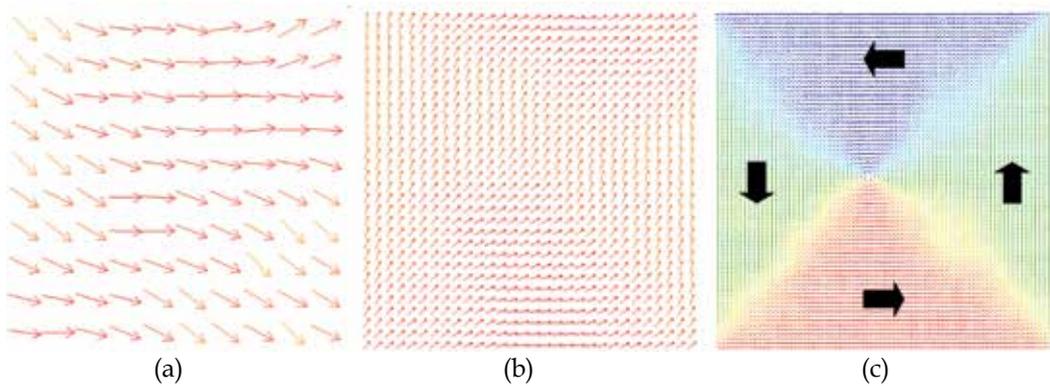


Fig. 6. Spin snapshots for different size square clusters with one side length of (a)  $L=10$ , (b)  $L=30$ , (c)  $L=75$  at lowest temperature. Closure domain structure of spin system appears for  $L=75$ . Arrows on (c) represent directions of magnetic domains.

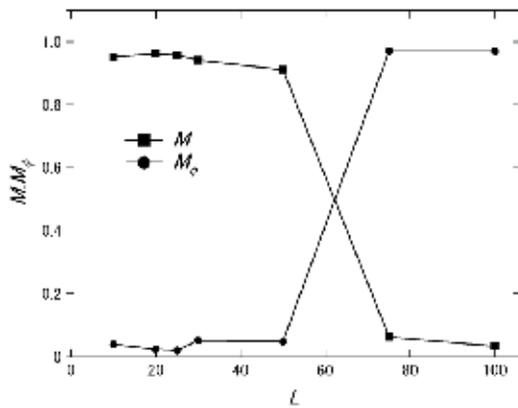


Fig. 7. Variation of  $M$  and  $M_\phi$  depending on square cluster size with  $L$ .

Here, magnetic susceptibilities of Europium chalcogenides were simulated as a function of temperature for a concrete example to demonstrate the usefulness of the parallelized MC program. Europium chalcogenides, such as EuO, EuS, EuSe, EuTe, are typical ionic magnetic materials (Mauger & Godart 1986). The crystal structure has NaCl type and two types of the exchange energy exist; that is,  $J_1$  for nearest site and  $J_2$  for second nearest site. These exchange energies change depending on the lattice constants. Magnetic properties show ferro-magnetism for  $|J_1| > |J_2|$  as EuO and antiferro-magnetism for  $|J_1| < |J_2|$  as EuTe.

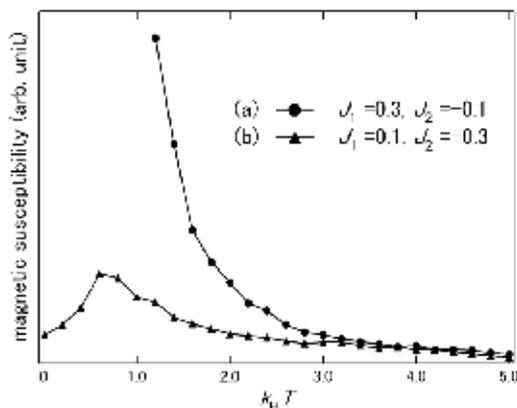


Fig. 8. Temperature dependence of magnetic susceptibilities of Europium chalcogenides for (a)  $J_1=0.3, J_2=-0.1$  and (b)  $J_1=0.1, J_2=-0.3$ .

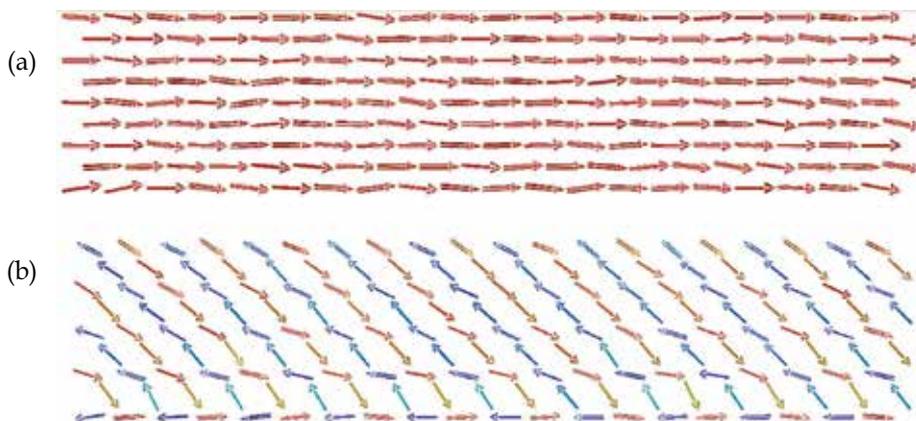


Fig. 9. Spin snapshots for a part of rectangular clusters of Eu chalcogenides with (a)  $J_1=0.3, J_2=-0.1$  and (b)  $J_1=0.1, J_2=-0.3$ .

Relative exchange energies were set as (a)  $J_1=0.3, J_2=-0.1$  and (b)  $J_1=0.1, J_2=-0.3$  for a rectangular cluster with each side length of  $5 \times 5 \times 50$ . These magnetic susceptibilities are estimated as gradients of the magnetization as a function of applied magnetic field  $B$  at each temperature. As shown in Fig. 8, the temperature dependence of magnetic susceptibilities has different behavior between (a) and (b). The susceptibility of (a) diverges around temperature  $k_B T=1.0$  and the magnetic property shows ferro-magnetism. The direction of the

magnetization aligns toward a longitudinal direction of the cuboids cluster by magnetic dipole interaction at low temperatures as shown in Fig. 9(a). The susceptibility of (b), on the other hand, has a peak around  $k_B T=0.8$  and the magnetic property shows antiferromagnetism. Their spins align as anti-parallel as shown in Fig. 9(b).

For large magnetic cluster with many spins, the parallelized MC method is very useful, although other MC method exists for huge clusters using FFT analysis (Sasaki & Matsubara 1997). The reason is that the parallelized MC method can directly deal with complicated interactions without any average operations, such as plural exchange interactions due to different elements or local interactions due to impurities and voids which are important for studying magnetic properties of real materials.

#### 4. Producing of magnetic domain

Magnetic domains in magnetic materials are produced by conflict among exchange interaction, magnetic dipole interaction and crystal anisotropy. In this section, using above MC method, the behavior of magnetic domains is represented. Here magnetic states were assumed that they depend on a Hamiltonian  $H$  including an exchange interaction energy  $H_J$ , a magnetic dipole interaction energy  $H_D$ , a magnetic anisotropy energy  $H_A$  and an applied magnetic field energy  $H_B$ ;

$$H = H_J + H_D + H_A + H_B. \quad (4)$$

$H_J$  term,  $H_D$  term and  $H_B$  term are same in Eq. (1).  $H_A$  term is given as following equations;

$$H_{A\_macro} = K_1 \sum_i \left( S_{i_x}^2 \cdot S_{i_y}^2 + S_{i_y}^2 \cdot S_{i_z}^2 + S_{i_z}^2 \cdot S_{i_x}^2 \right), \quad (5a)$$

$$H_{A\_micro} = A \sum_i \left( \frac{1}{|\mathbf{r}_{ij} - a_r \mathbf{S}_i|} - \frac{1}{|\mathbf{r}_{ij}|} \right). \quad (5b)$$

Equation (5a) is usual anisotropy representation for bcc crystal structure and Eq.(5b) is microscopic conventional anisotropy which was introduced to study for a deformed cluster. Below the parameters were set to  $J_{ij}=1.0$ ,  $D=0.1$ ,  $K_1=1.0$ ,  $A=5$  and  $a_r=0.3$ , respectively. These are tentative values to examine the usefulness of the model. The effective radius was set to  $r_{eff}=0.97$  when excluding the second nearest neighbor spins in bcc structure.

Two spin systems of bcc structure with the lattice constant  $L=1$  were formed into a cylindrical cluster with a diameter of  $28L$  and  $2L$  thickness including the number of 3291 spins and a spherical cluster with a diameter of  $18L$  including the number of 7239 spins.

Figure 10 shows the temperature dependence of the closure domain structure parameter  $M_\phi$  for the cylindrical cluster using each Hamiltonian; (a)  $H_J+H_D$ , (b)  $H_J+H_D+H_{A\_macro}$ , (c)  $H_J+H_D+H_{A\_micro}$ . Here  $M_\phi$  is defined as same as Eq.(3);

Note that  $M_\phi$  at the lowest temperature appears to be in the stable state, because it is the result after cooling down from sufficiently higher temperatures. Then the result without any anisotropies (a) shows  $M_\phi=1.0$ , on the other hand, ones with anisotropies (b) and (c) show  $M_\phi=0.95$ . The decreases of  $M_\phi$  for the calculations with both anisotropies are due to producing magnetic domain walls (DWs). As shown in Fig.11(b), four divided magnetic domains were produced with 90 degree DWs (Neel walls); almost the spins align toward the

x-axis [100] and the y-axis [010], nevertheless the spin directions gradually change in Fig.11(a). When using  $H_J+H_D+H_{A\_micro}$  the snapshot at the lowest temperature shows almost similar to Fig.11(b). As shown in Fig. 11, the effect of  $H_A$  is reflected in magnetic domain producing on a cylindrical cluster.

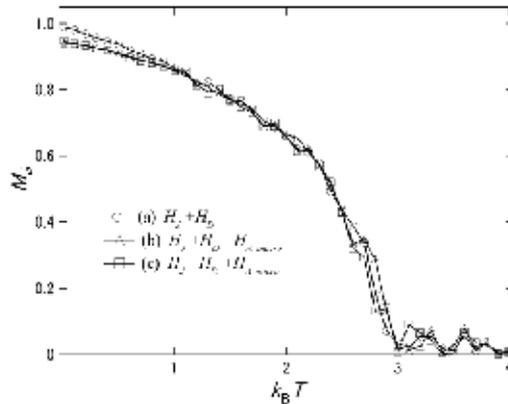


Fig. 10. Closure domain parameter  $M_\phi$  as a function of temperature  $k_B T$  for a cylindrical cluster using Hamiltonian; (a)  $H_J+H_D$ , (b)  $H_J+H_D+H_{A\_macro}$ , (c)  $H_J+H_D+H_{A\_micro}$ .

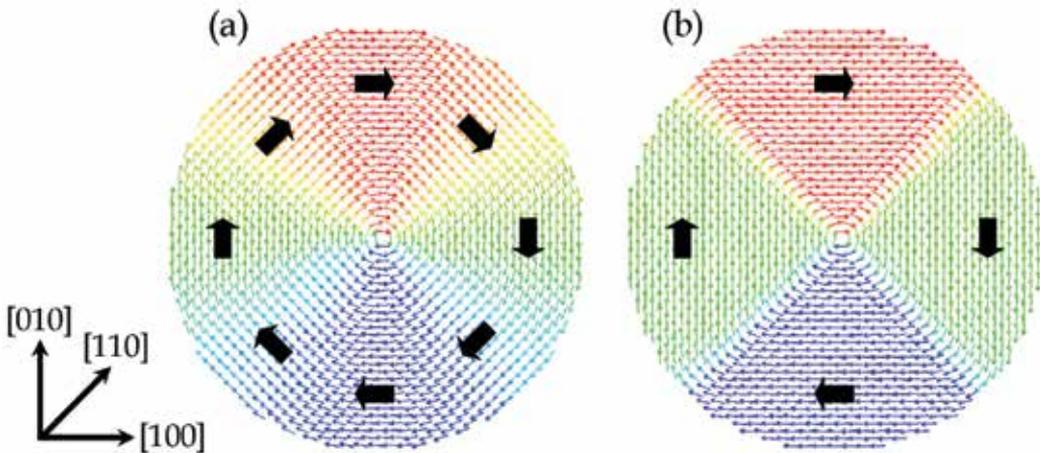


Fig. 11. Spin snapshots for a cylindrical cluster at the lowest temperature using Hamiltonian; (a)  $H_J+H_D$ , (b)  $H_J+H_D+H_{A\_macro}$ .

Figure 11 shows the effect of  $H_A$  for magnetic domain producing in a cylindrical cluster. As shown in Fig. 11(b), four divided magnetic domains were produced with 90 degree domain walls (Neel walls); almost the spins align toward the x-axis [100] and the y-axis [010], nevertheless the spin directions gradually change in Fig. 11(a). When using  $H_J+H_D+H_{A\_micro}$  the snapshot at the lowest temperature shows almost similar to Fig. 11(b).

Figure 12 shows magnetizations as a function of applied magnetic field (M-H curves) at the temperature of  $k_B T=0.1$  along the [100] and [110] directions for the cylindrical cluster using  $H_1=H_J+H_D+H_{A\_macro}+H_B$  including the macroscopic anisotropy and  $H_2=H_J+H_D+H_{A\_micro}+H_B$

including the microscopic anisotropy. For both Hamiltonians, the anisotropy properties correspond qualitatively to the experimental result of bcc iron's one; the M-H curves show the magnetization along the [100] direction rapidly increases and reaches the saturated magnetization soon, and one along the [110] direction increases slowly on the way, therefore the [100] direction is the axis of easy magnetization for the cluster (Kittel 1986).

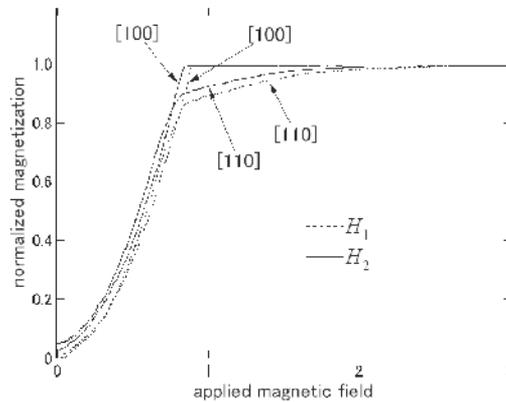


Fig. 12. Magnetizations as a function of applied magnetic field along the [100] and [110] directions for a cylindrical cluster using  $H_1 = H_J + H_D + H_{A\_macro} + H_B$  and  $H_2 = H_J + H_D + H_{A\_micro} + H_B$ .

Figure 13 shows spin snapshots on the magnetization processes for the cylindrical cluster using  $H_2$ , when the magnetic field was applied along the [100] direction and the [110] direction. For the magnetic field along the [100] direction, DWs are monotonously moving and the magnetic domain including the spins toward the [100] direction in four divided magnetic domains gradually grow with increasing the magnetic field up to the saturation magnetization around  $B=0.85$ . On the other hand, for the magnetic field along the [110] direction, at first, two magnetic domains including the spins toward the [100] and the [010] directions grow and form one big DW at around  $B=0.85$ . Then the DW was fixed and the spins in the two domains gradually rotate toward the [110] direction, that is, rotation magnetization. In Fig.12, the slope of the M-H curve with the applied magnetic field along the [110] direction decreases more than around  $B=0.8$  and the result depends on the slow reaction of the rotation magnetization with increasing magnetic fields.

Figure 14 shows M-H curves at the temperature of  $k_B T = 0.1$  along the [100], [110] and [111] directions for the spherical cluster using  $H_1$  and  $H_2$ . The results show the [111] direction is the axis of hard magnetization as similar as the experimental results of bcc iron (Kittel 1986). Above magnetic properties using  $H_2$  as shown in Fig. 12, Fig. 13 and Fig. 14 well correspond to the results of the simulation using  $H_1$ . As a result, it would be possible to deal with  $H_2$  as alternative to  $H_1$ . An advantage of  $H_2$  including the microscopic anisotropy is to simulate magnetic processes for deformed clusters which have local crystal asymmetry.

Figure 15 shows spin snapshots on the magnetization processes for the original cylindrical cluster and the cylindrical cluster elongated 1.01 times along the [010] direction as a deformed cluster using  $H_2$ , when the magnetic field was applied along the [110] direction. Here the parameter  $A$  in (5b) is set to  $A=10$  for more clearly checking the effect of the anisotropy. The results for the original cluster (left side in Fig. 15) are similar to ones in Fig.13 (right side). But the results for the deformed cluster, after the big DW produced by

the growth of two magnetic domains, the DW is still moving with rotation magnetization more than  $B=0.85$ , that is, the DWD has two steps process. The latter DWD would be regarded as the balance of pressure on DW broke due to asymmetric anisotropy in terms of "equation of motion for DW". But above model can introduce the DWD behavior naturally without importing other parameters.

The difference of the DWD behavior between the original cylindrical cluster and the deformed cluster does not clearly affect the M-H curves as shown in Fig. 16. This means the measurements of M-H curves could not give any efficient information for DWD. Then the other measurement such as Barkhausen noise would be needed to more exactly know DWD behavior.

As mentioned above, MC simulations using  $H_2$  including a microscopic anisotropy will be useful to study for DWD behavior, although now the results correspond to experimental one only qualitatively.  $H_{A\_micro}$  in  $H_2$  is originally introduced as crystal field from surrounding ligands, that is, a summation of Coulomb potentials. In general the charges in metals are strongly screened by conduction electrons. Therefore  $H_{A\_micro}$  should be rather thought as a representation of a hybridization effect between electron wave functions, then the parameter  $A$  and  $a_r$  in  $H_{A\_micro}$  would concern with the intensity of transfer integrals and the effective radius of the wave function respectively. As a result, the proposed model has a possibility to connect DWD behavior with material properties more deeply.

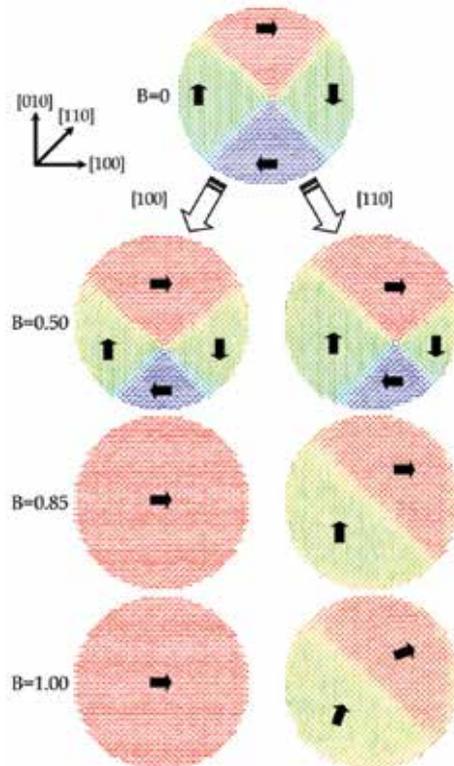


Fig. 13. Spin snapshots on magnetization processes for a cylindrical cluster using  $H_2 = H_J + H_D + H_{A\_micro} + H_B$ , when magnetic fields were applied along the [100] direction (left side) and along the [110] direction (right side).

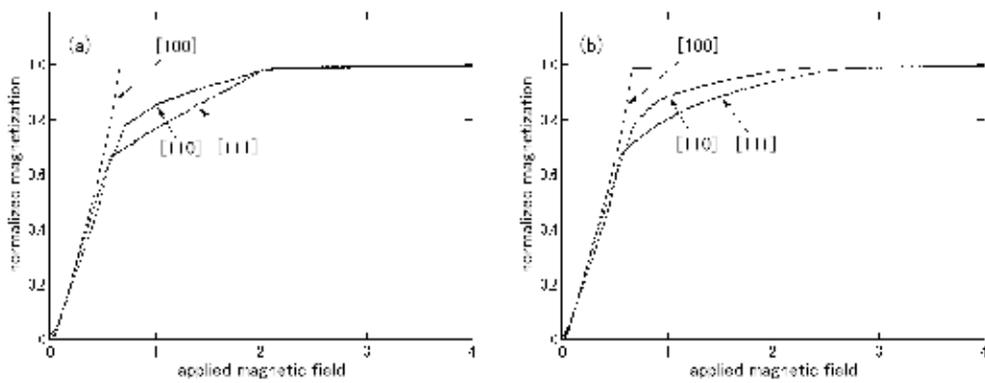


Fig. 14. Magnetization as a function of applied magnetic field along the [100], [110] and [111] directions for a spherical cluster using (a)  $H1=H_J+H_D+H_{A\_macro}+H_B$  and (b)  $H2=H_J+H_D+H_{A\_micro}+H_B$ .

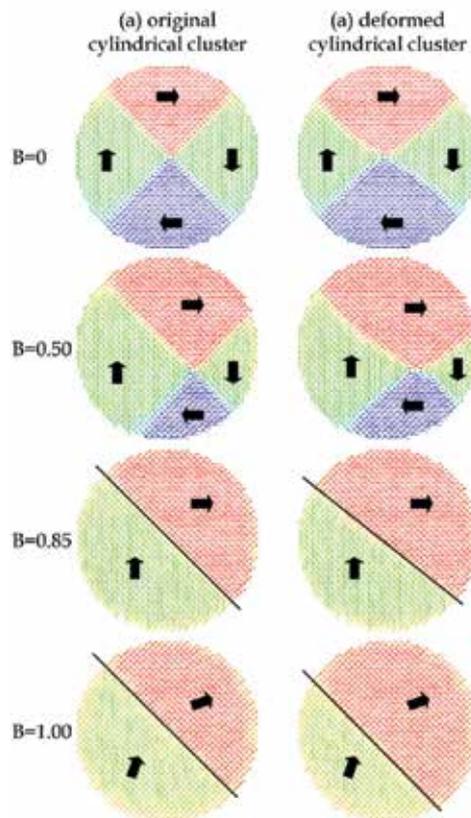


Fig. 15. Spin snapshots on magnetization processes for (a) the original cylindrical cluster and (b) the deformed cylindrical cluster elongated 1.01 times along the [010] direction, when magnetic fields were applied along the [110] direction. Note that parameter  $A$  in (4b) is set to  $A=10$ .

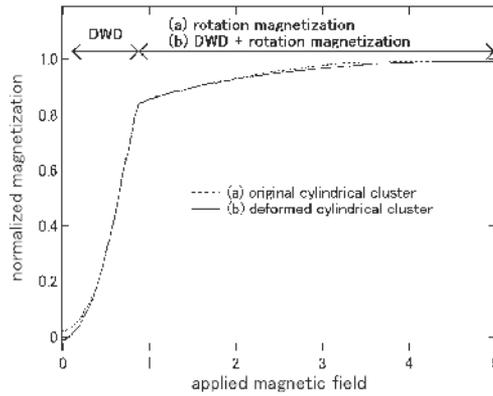


Fig. 16. Magnetizations as a function of applied magnetic fields along the [110] direction for (a) the original cylindrical cluster and (b) the deformed cylindrical cluster using  $H_2 = H_J + H_D + H_{A\_micro} + H_B$ . Note that parameter  $A$  in (4b) is set to  $A = 10$ .

## 5. DWD for nano-wire

In this section, based on above method, the behavior of magnetic domain wall displacement (DWD) for nano-wire is simulated, which is important study for spintronics (Yamaguchi *et al.* 2009).

Here a rectangular solids spin system composed of  $5 \times 5 \times 150$  cells ( $0 \leq x \leq 4$ ,  $0 \leq y \leq 4$ ,  $0 \leq z \leq 149$ ) standing for a nano-wire was prepared as a normal spin system without any local disorder. A following Hamiltonian was used for the simulation:

$$\begin{aligned}
 H &= H_J + H_D + H_B \\
 &= - \sum_{near} J_{ij} \mathbf{S}_i \cdot \mathbf{S}_j + D \sum_{all} \left( \frac{\mathbf{S}_i \cdot \mathbf{S}_j}{|\mathbf{r}_{ij}|^3} - \frac{3}{|\mathbf{r}_{ij}|^5} (\mathbf{S}_i \cdot \mathbf{r}_{ij})(\mathbf{S}_j \cdot \mathbf{r}_{ij}) \right) + B \sum_i \mathbf{S}_i.
 \end{aligned} \quad (1)$$

In this simulation, the parameters were set as  $J_{ij} = 1.0$  between normal spins,  $r_{eff} = 1.0$ ,  $D = 0.1$ . The value of  $\mathbf{S}_i$  was fixed as  $|\mathbf{S}_i| = 1$ . In this section, for simplicity, above Hamiltonian has no crystal anisotropy, although it has an important role for producing magnetic domains as shown in section 4. Here, alternatively, a shape magnetic anisotropy due to magnetic dipole interaction between spins produces magnetic domains.

Figure 17 shows temperature dependence of normalized magnetization  $M$  gradually cooling down from  $k_B T = 2.0$  to  $k_B T = 0.01$  for the rectangular cluster whose initial spin states were taken as random directions. Here  $M$  is defined as below

$$M = \frac{1}{N} \left| \sum_i \mathbf{S}_i \right|. \quad (6)$$

At each temperature,  $M$  is determined after  $N$  MCS repeating for producing the results in equilibrium. The curve obeys the Curie Weiss law and it has the Curie temperature of about  $k_B T_c = 1.5$ . At the lowest temperature, almost spins align toward the longitudinal direction of the rectangular cluster due to the shape magnetic anisotropy as shown in Fig.18. Figure 19 shows applied magnetic field dependence of normalized magnetization  $M_z$ , that is,

magnetic hysteresis curve. The direction of magnetic field  $B$  is set to the axis of  $z$  and applied on the process  $B = 0 \rightarrow +1.0 \rightarrow -1.0 \rightarrow +1.0$  with the step width  $\Delta B = 0.01$ . Here,  $M_z$  is defined as below

$$M_z = \frac{1}{N} \sum_i \mathbf{S}_i \cdot \mathbf{k}. \quad (7)$$

Here,  $\mathbf{k}$  is the unit vector along  $z$ -axis. The rectangular cluster has a large coercive force which would be due to the shape magnetic anisotropy.  $M_z$  is saturated under the magnetic field of  $B=0.5$ .

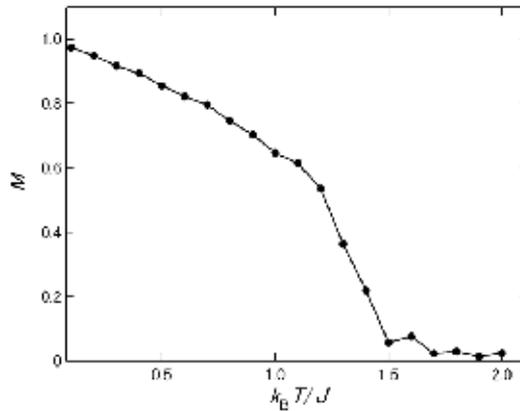


Fig. 17. Temperature dependence of normalized magnetization  $M$  for the rectangular cluster composed of  $5 \times 5 \times 150$  spins.  $M$  was simulated cooling down from higher temperatures.

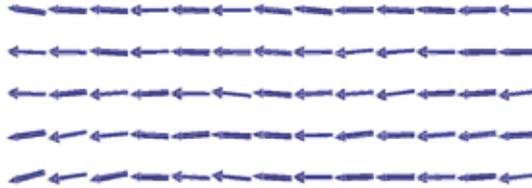


Fig. 18. Snapshot of the spin structure for the left edge of the rectangular cluster at the lowest temperature.

Next the constant reversal magnetic field of  $B = +0.5$  was applied for the rectangular cluster with  $M_z = -1.0$  at the lowest temperature in Fig. 17. Figure 20 shows the time dependence of  $M_z$  until 20000 MCS. The changing of  $M_z$  is small until 2500 MCS, and  $M_z$  changes with the almost constant gradient from 2500 MCS to 10000 MCS. Then  $M_z$  becomes constant over 10000 MCS, that is, saturation magnetization. The period until 2500 MCS is an initial step of the reversal magnetization process that spin directions were first reversed from sites around both longitudinal edge sides ( $z=0$  and  $z=149$ ) but obvious DWs are not produced yet. In the second period between 2500 and 10000 MCS, double DWs are produced around double edges of the rectangular cluster, as shown in Fig. 21(a), which shows a snapshot of the spin structure at  $t=3000$  MCS. In the snapshot, there are double DWs at around  $z=10$  and  $z=140$  and the spins in the DWs take a screw structure, don't take Bloch or Neel typed DWs, as

shown in Fig. 21(b). Spin snap shots are shown in Fig. 21(c) on each MCS; 0 MCS, 3000 MCS, 6000 MCS and 10000 MCS.

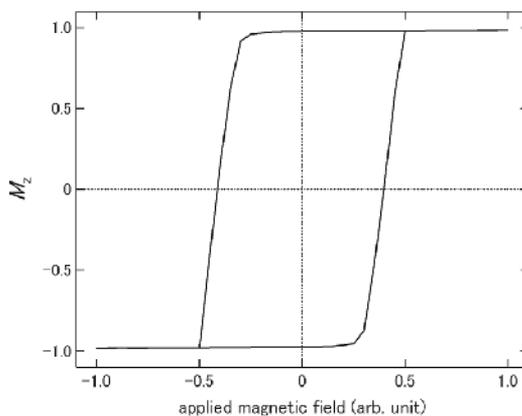


Fig. 19. Applied magnetic field dependence of  $M_z$  (hysteresis curve) for the rectangular cluster with  $M_z = -1.0$  at initial state.

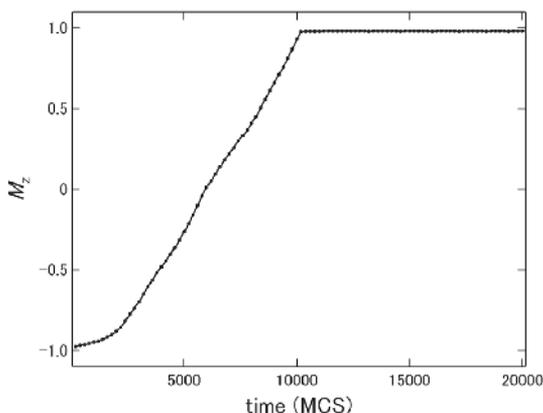


Fig. 20. Time dependence of  $M_z$  under the constant reversal magnetic field of  $B = +0.5$  for the rectangular cluster with  $M_z = -1.0$ .

These DWs run toward the middle of the cluster until 10000 MCS as shown in Fig. 22. In this figure, each line shows an average of absolute value of the  $z$  component of spins ( $=S_z$ ) included on the  $x$ - $y$  plane at each  $z$  position at each increasing time elapse. Then each dip on line corresponds to the DW position, because  $S_z$  becomes smaller around DW than ones in other positions. In the last step, the double DWs vanish after encounter each other around the middle of the rectangular cluster over 10000 MCS.

Figure 23 shows the DW position depending on time elapses. In this model, using gradients of the DW position line for time, the DW velocity was estimated as  $0.93 \times 10^{-2}$  (cell/MCS) for the rectangular cluster without impurities. Note that the velocity cannot be estimated by  $M_z$  in Fig. 19, because the rectangular cluster has double DW on reversal magnetization process and the increasing of  $M_z$  is the result that the effects of double DWs are superposed.

Here local disorders by magnetic impurities are introduced into the rectangular cluster as a normal spin system. These local disorders are randomly spread over the rectangular cluster until the number corresponding to the densities. Introducing of magnetic impurities is supposed to change no parameters of normal spins except for exchange interaction  $J_{ij}$ . The exchange interactions is set as  $J_{ij} = 1.5$  between a normal spin and an impurity, and  $J_{ij} = 2.0$  between impurities expecting magnetic enhancement due to the impurity.

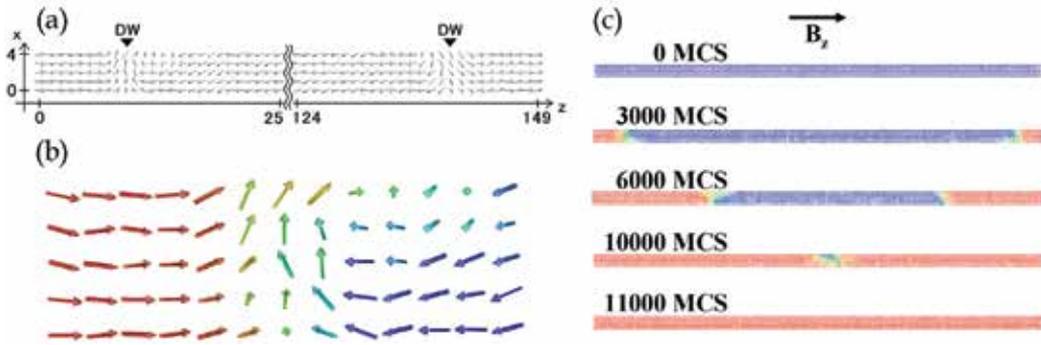


Fig. 21. (a) Snapshot of the spin structure during reversal magnetic field for the rectangular cluster at  $t=3000$  MCS after the magnetic field was applied. (b) Enlarged view of snapshot of the spin structure around the left side DW in (a). (c) Spin snapshots on each MCS; 0 MCS, 3000 MCS, 6000 MCS and 10000 MCS.

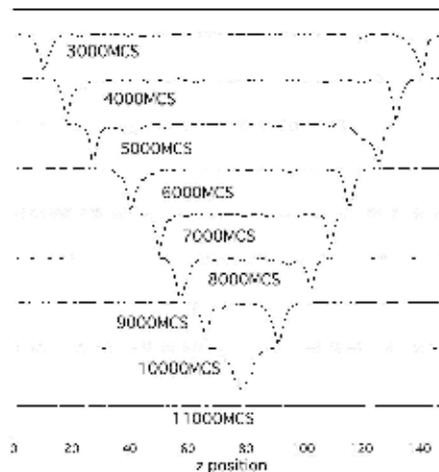


Fig. 22. Average of absolute value of  $S_z$  at each  $z$  position at each increasing time elapse, respectively. Each dip shows the DW position.

Figure 24 shows time dependence of DW position changes ( $\Delta DWD$ ) for the rectangular cluster with magnetic impurities, since obvious DW is produced under the reversal magnetic field. It is clearly seen that the gradients decrease with increasing the density of impurities.

Figure 25 shows variations of DWD velocity depending on impurities density. DWD velocity was found to decrease with increasing impurity.

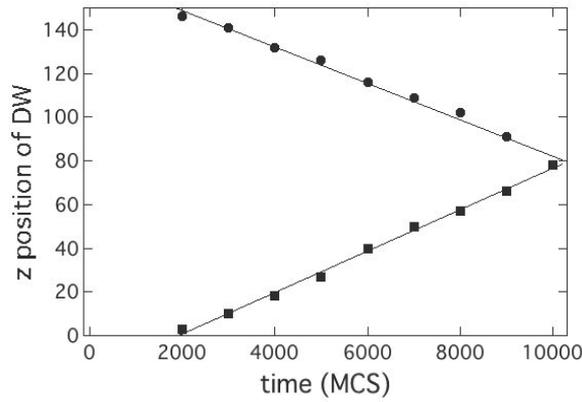


Fig. 23. Time dependence of the DW position on the left side (square markers) and right side (circle markers) of the rectangular cluster.

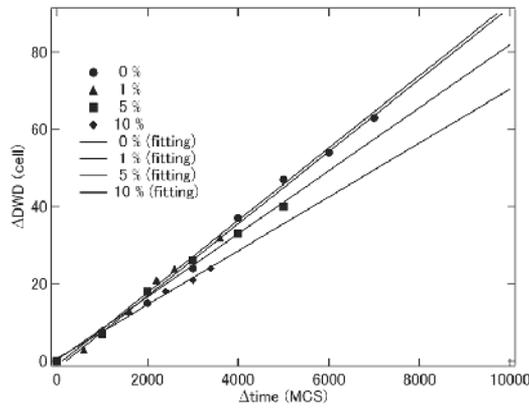


Fig. 24. Time dependence of the DW position changes of the rectangular cluster with various densities of magnetic impurities.

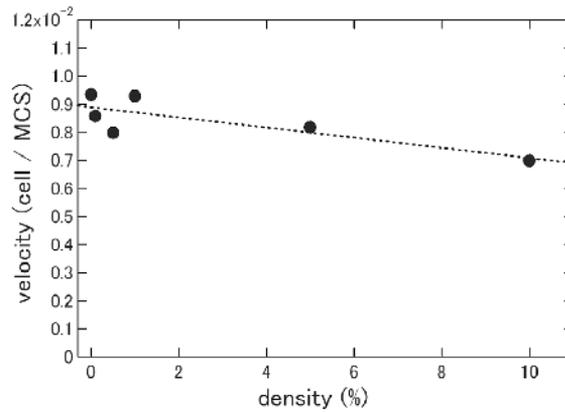


Fig. 25. DW velocity changes with magnetic impurity densities.

In this section, DWD velocities were estimated for the rectangular clusters with different densities of magnetic impurities by MC simulation. The method above mentioned for investigating the behavior of DW will be useful for the development of nano-magnetic devices in near future.

## 6. M-H curves with a local magnetic distribution

The nickel-base superalloy Alloy 600 (Inconel) is widely used as structural materials for their high mechanical strength, e.g. for atomic power plants, and therefore early detection of the fatigue of the materials is very important. It is known that the sensitization of Alloy 600 due to chromium (Cr) depletion near the grain boundary by thermal heat treatment causes the intergranular stress corrosion cracking (IGSCC), then especially the behavior under the sensitization has been studied as pressing matters (Kowaka *et al.* 1981, Wang & Gan 2001, Mayo 2004). It has been also known that the sensitization produced the magnetism in Alloy 600 which has no magnetism originally (Aspden *et al.* 1972, Takahashi *et al.* 2004b). Recently relationship between magnetic properties and sensitization is focused with expectation for potentiality of nondestructive evaluation (NDE) (Takahashi *et al.* 2004a). Several experimental reports show the magnetization occurs at Cr depletion areas around grain boundaries and the degree of sensitization affects the magnetic properties such as magnetic hysteresis (M-H) curves. But now it is not solved yet how the distribution of Cr depletion affects the change of magnetism in Alloy 600, although the relationship between the distribution of Cr depletion and the magnetism is important to estimate of the degree of sensitization using magnetic NDE.

In this section, magnetic properties of sensitized Alloy 600 by different heating duration times were simulated using Monte Carlo (MC) method and the results are discussed focusing on M-H curves affected by the sensitization (Yamaguchi *et al.* to be published).

A cubic system composed of  $31^3$  cells ( $0 \leq x \leq 30$ ,  $0 \leq y \leq 30$ ,  $0 \leq z \leq 30$ ) was prepared including magnetic sites with a distribution. The distribution was decided by Cr depletion degree around a grain boundary on the supposition that Cr depletion introduces magnetic moments around the depletion area (Aspden *et al.* 1972, Takahashi *et al.* 2004b). The distribution of Cr depletion depending on heating duration time was calculated by thermodynamic analysis (Pruthi *et al.* 1977, Was & Kruger 1985, Grujicic & Tangrila 1991, Kai *et al.* 1993, Bao *et al.* 2006). Here the heating duration time means the period of thermal annealing under a constant heating temperature. Figure 26 shows the calculation results of the distributions of Cr depletion with each duration time (1h, 25h, 50h, 150h) under the heating temperature at 650 Celsius degree. The distributions of magnetic sites along x-axis of the cubic system corresponding to the distribution of Cr depletion are shown in Fig.27 as the surface view of the clusters. Here red circles represent the magnetic sites produced with a probability obeying the distribution of Cr depletion and blue circles are non magnetic sites. In Fig.27, the grain boundary is set on the y-z plane at the x-coordination of 15 and the edge surface coordination  $x=0$  and  $x=30$  are regarded as -300nm and +300nm in Fig.26 respectively.

A following Hamiltonian was used for the simulation:

$$\begin{aligned}
 H &= H_J + H_D + H_B \\
 &= -\sum_{near} J_{ij} \mathbf{S}_i \cdot \mathbf{S}_j + D \sum_{all} \left( \frac{\mathbf{S}_i \cdot \mathbf{S}_j}{|\mathbf{r}_{ij}|^3} - \frac{3}{|\mathbf{r}_{ij}|^5} (\mathbf{S}_i \cdot \mathbf{r}_{ij})(\mathbf{S}_j \cdot \mathbf{r}_{ij}) \right) + B \sum_i \mathbf{S}_i.
 \end{aligned} \tag{1}$$

In this simulation, the parameters were set as  $J_{ij}=1.0$ ,  $r_{eff}=1.0$ ,  $D=0.01$  in Eq.(1).

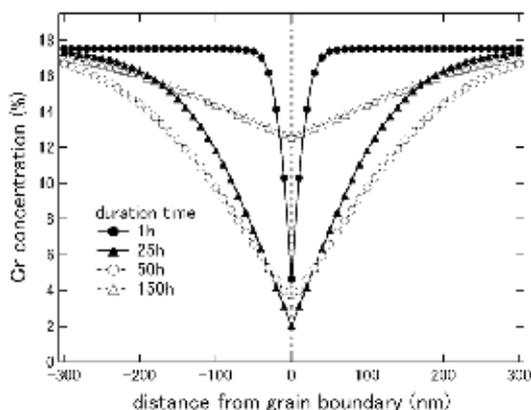


Fig. 26. Distribution of Cr depletion as a function of distance from grain boundary for each heating duration time.

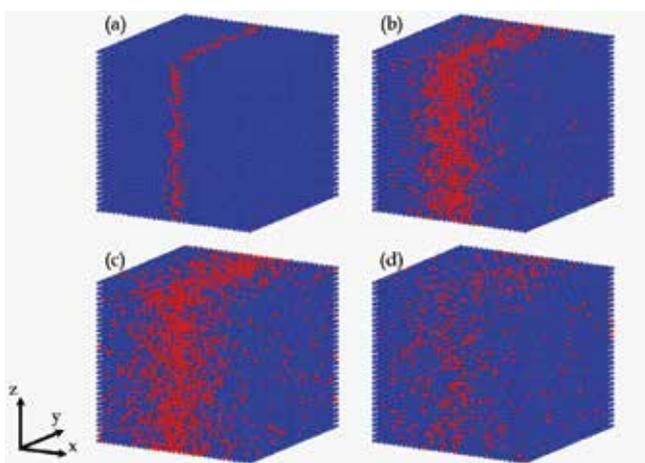


Fig. 27. Surface view of model clusters including magnetic sites due to the distribution of Cr depletion for duration time of (a) 1h, (b) 25h, (c) 50h and (d) 150h. Red circles and blue circles represent magnetic sites and non magnetic sites, respectively.

Figure 28(a) shows the experimental results of the magnetic M-H curves for Alloy 600 with different heating duration times. The measurements were performed at room temperature using vibration sample magnetometer (VSM). On the other hand, Fig. 28(b) shows the calculation results of M-H curves. The results of calculated M-H curves are the average of magnetization for two directions of applied magnetic field along perpendicular (x direction) and parallel (y direction) to grain boundary surface of cubic system, and the magnetization are normalized by total cell number ( $=31^3$ ). The applied magnetic field in this calculation is represented as arbitrary unit, and the value of 0.2 roughly corresponds to 2000 A/m in experiment from the estimation of magnetic field for saturation magnetization of the cluster with duration time of 50h. The behaviors of calculated M-H curves for duration times

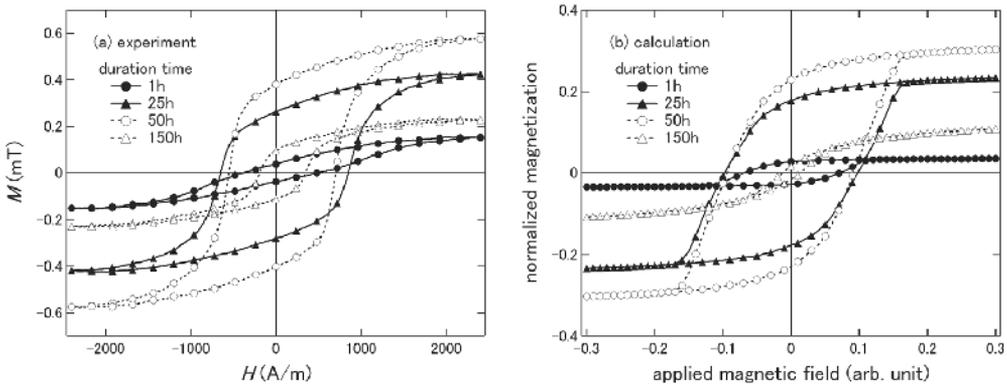


Fig. 28. M-H curves of (a) experiment and (b) calculation for each duration time.

correspond to the experimental ones, especially for the residual magnetization  $M_r$  and magnetic coercivity  $H_c$  which are important values in the demagnetizing curve. Figure 29(a) and 29(b) show the heating duration time dependence of  $M_r$  and  $H_c$  respectively, including more different duration times. The calculation result (solid line) has good correspondence with the experimental ones (dashed line). The difference of the duration time at  $M_r$  maximum between calculation and experiment can be due to the reliability of the estimated distribution of Cr depletion in Fig.26.

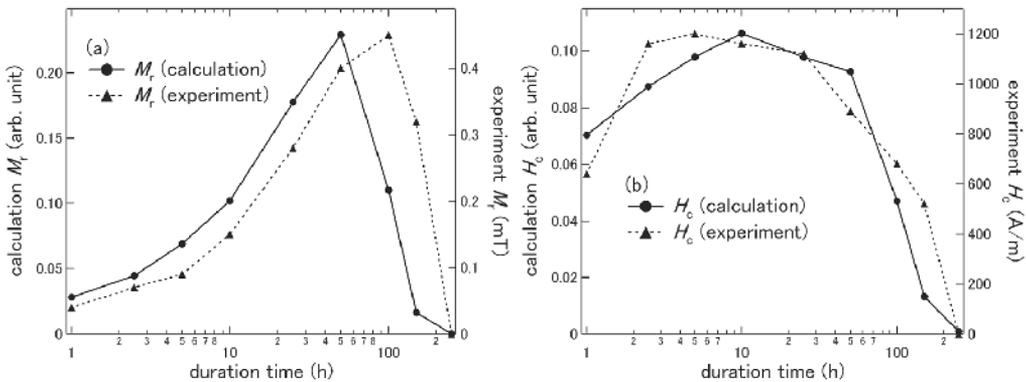


Fig. 29. Duration time dependence of (a)  $M_r$  and (b)  $H_c$  for experiment and calculation results.

To discuss focusing on the relationship between the distribution of magnetic site (= Cr depletion) and magnetic properties, such as  $M_r$  and  $H_c$ , the number of total magnetic sites in the cubic system and the average number of nearest neighbor magnetic sites are shown in Fig. 30(a) and 30(b), respectively as a function of the duration time. Here note the number of nearest neighbor magnetic sites for each magnetic site can range between 0 and 6, therefore the average number of nearest neighbor magnetic sites is different for each cluster corresponding to the distribution of Cr depletion as shown in Fig. 27. As shown in Fig. 29(a) and 30(a),  $M_r$  obeys the number of total magnetic sites. The result is reasonable in the view point that  $M_r$  is almost proportionate to the saturation magnetization. On the other hand,  $H_c$  nearly corresponds to the average number of nearest neighbor magnetic sites as shown in Fig. 29(b) and 30(b). In other words,  $H_c$  is affected by the density of magnetic sites around

grain boundary. Hence, these results suggest that the distribution of Cr depletion by sensitization, that is, the total amount of Cr depletion and the density of Cr depletion around grain boundaries can be estimated by  $M_r$  and  $H_c$ , respectively.

Above calculation model uses the exchange interaction with effective radius  $r_{eff} = 1.0$ , then the effective strength of the exchange interaction depends on the number of the nearest neighbor magnetic sites. Now let us see the behavior of the effective interaction depending on the duration time in the view point of Curie temperature  $T_c$  which depends on the exchange interaction.

Figure 31 shows the temperature dependence of calculated magnetization without applied magnetic field for different duration times. The temperature below which the spontaneous magnetization appears, that is,  $T_c$  is depending on each duration time. To estimate  $T_c$  more exactly, temperature dependence of magnetic susceptibility  $\chi$  is calculated by M-H curve for each temperature such as shown in Fig. 32. Figure 33 shows the temperature dependence of  $1/\chi$  and  $T_c$  is estimated as the cross point of the temperature axis. Then the duration time dependence of  $T_c$  is also following the average number of the nearest magnetic sites as shown in Fig.34. The result suggests the effective exchange interaction affects both  $H_c$  and  $T_c$  through the density of magnetic sites due to Cr depletion around grain boundaries.

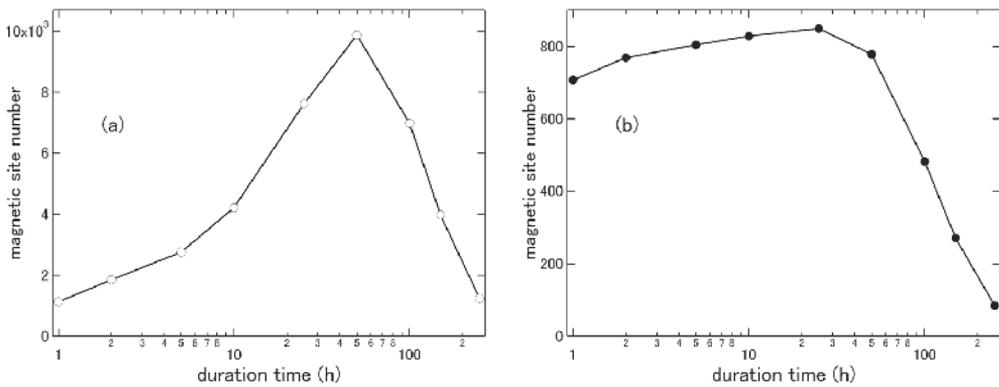


Fig. 30. Duration time dependence of (a) number of total magnetic sites and (b) average number of the nearest neighbor magnetic sites.

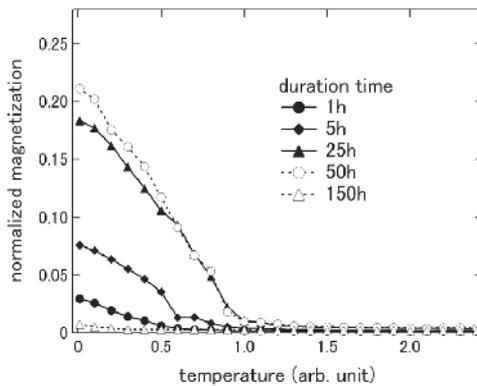


Fig. 31. Temperature dependence of calculated magnetization for each duration time

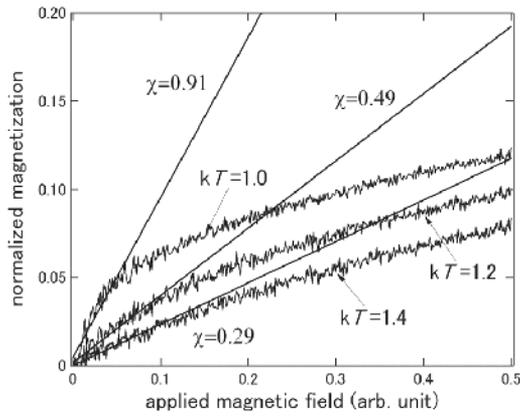


Fig. 32. Example of calculation for magnetic susceptibility  $\chi$  from M-H curve at each temperature for the duration time of 25h.

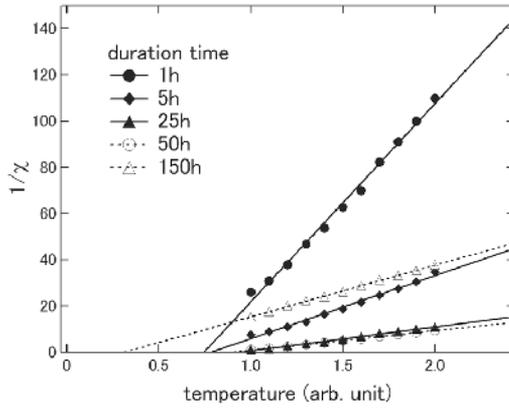


Fig. 33. Temperature dependence of inverse of calculated magnetic susceptibility  $1/\chi$  for each duration time.

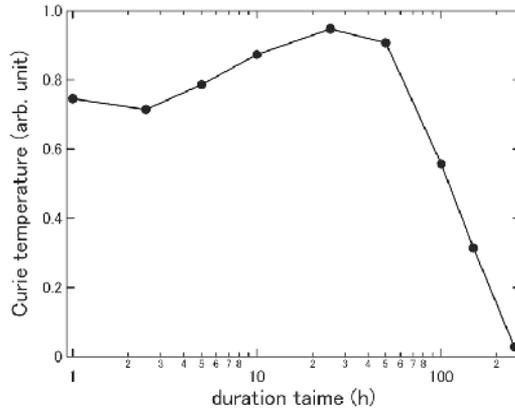


Fig. 34. Duration time dependence of Curie temperature.

In above the model, magnetic particles due to Cr depletion disperse around a grain boundary in Alloy 600 and it can be regarded as a magnetic granular structure with a distribution. Then  $M_r$  and  $H_c$  on a M-H curve tell the total amount and the density of Cr depletion around grain boundaries, respectively. Therefore the analysis of magnetic dynamic process using Monte Carlo method would tell the degree of sensitization due to fatigue for Alloy 600.

## 7. References

- Aspden, R. G.; Economy, G.; Pement, F. W. & Wilson, I. L. (1972). Relationship Between Magnetic Properties, Sensitization, and Corrosion of Incoloy Alloy 800 and Inconel Alloy 600. *Metallurgical Transactions*, Vol. 3, 2691-2697
- Bao, G.; Shinozaki, K.; Inkyo, M.; Miyoshi, T.; Yamamoto, M.; Mahara, Y. & Watanabe, H. (2006). Modeling of precipitation and Cr depletion profiles of Inconel 600 during heat treatments and LSM procedure. *Journal of Alloys and Compounds*, Vol. 419, No.1-2, August, 118-125
- Grujicic, M. & Tangrila, S. (1991). Thermodynamic and kinetic analyses of time-temperature-sensitization diagrams in austenitic stainless steels. *Materials Science and Engineering*, Vol.A142, No.2, August, 255-259
- Kai, J. J.; Tsai, C. H. & Yu, G. P. (1993). The IGSCC, sensitization, and microstructure study of Alloys 600 and 690\*. *Nuclear Engineering and Design*, vol. 144, No.3, November, 449-457
- Kittel, C. (1986). *Introduction to Solid State Physics, 6th ed.*, John Wiley & Sons, Inc., ISBN, New York
- Kowaka, M.; Nagano, H.; Kudo, T. & Okada, Y. (1981). Effect of Heat Treatment on The Susceptibility To Stress Corrosion Cracking of Alloy 600. *Nuclear Technology*, Vol. 55, 394-404
- Landau, D. P. & Binder, K. (2000). *A Guide to Monte Carlo Simulations in Statistical Physics*, Cambridge University Press, 0521653665, Cambridge
- Mauger, A. & Godart, C. (1986). The magnetic, optical, and transport properties of representatives of a class of magnetic semiconductors: The Europium chalcogenides. *Phys. Rep.*, Vol. 141, No.2-3, 51-176
- Mayo, W. E. (2004). Predicting IGSCC/IGA susceptibility of Ni-Cr-Fe alloys by modeling of grain boundary chromium depletion. *Materials Science and Engineering A*, Vol.232, No.1-2, 129-139
- Metropolis, N.; Rosenbluth, A.; Rosenbluth, M. & Teller, A. (1953). Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.*, Vol.21, No.6, 1087-1092
- Pruthi, D. D.; Anand, M. S. & Agarwala, R. P. (1977). Diffusion of Chromium in Inconel-600. *Journal of Nuclear Material*, Vol. 64, No.1-2, January, 206-210
- Sasaki, J. & Matsubara, F. (1997). Circular phase of a two-dimensional ferromagnet with dipolar interaction. *J. Phys. Soc. Jpn*, Vol.66, No.7, 2138-2146, 00319015
- Takahashi, S.; Sato, H.; Kamada, Y.; Ara, K. & Kikuchi, H. (2004a). A new magnetic NDE method in inconel 600 alloy, IOS Press, Vol. 19, 3-8
- Takahashi, S.; Sato, Y.; Kamada, Y. & Abe, T. (2004b). Study of chromium depletion by magnetic method in Ni-based alloys. *Journal of Magnetism and Magnetic Materials*, Vol. 269, 139-149

- Vedmedenko, E. Y.; Oepen, H. P.; Ghazali, A.; Levy, J. C. S. & Kirschner, J. (2000). Magnetic Microstructure of the Spin Reorientation Transition: Computer Experiment. *Phys.Rev.Lett.*, Vol.84, No.25, 5884-5887
- Wang, J. D. & Gan, D. (2001). Effects of grain boundary carbides on the mechanical properties of Inconel 600. *Materials Chemistry and Physics*, Vol. 70, No.2, 124-128
- Was, G. S. & Kruger, R. M. (1985). A thermodynamic and kinetic basis for understanding chromium depletion in Ni-Cr-Fe alloys. *Acta Metallurgica*, Vol. 33, No.5, May, 841-854
- Yamaguchi, K.; Tanaka, S.; Nittono, O.; Takagi, T. & Yamada, K. (2004). Monte Carlo simulation of dynamic magnetic processes for spin system with local defects. *Physica B*, Vol. 343, No.1-4, January, 298-302
- Yamaguchi, K.; Suzuki, K.; Nittono, O.; Yamada, K.; Enokizono, M. & Takagi, T. (2009). Monte Carlo Simulation for Magnetic Domain Wall Displacements in Magnetic Nano-Wires with Local Disorders. *IEEE Trans. Magn.*, Vol. 45, No.3, March, 1622-1625
- Yamaguchi, K.; Suzuki, K.; Nittono,; Uchimoto, T. & Takagi, T. (to be published). Magnetic Dynamic Process of Magnetic Layers around Grain Boundary for Sensitized Alloy 600. *IEEE Trans. Magn.*.

# Monte Carlo Simulations of Grain Growth in Polycrystalline Materials Using Potts Model

Miroslav Morháč<sup>1</sup> and Eva Morháčová<sup>2</sup>

<sup>1</sup>*Institute of Physics, Slovak Academy of Sciences,  
Dubravska cesta 9, 845 11 Bratislava,*

<sup>2</sup>*Faculty of Mechanical Engineering, Slovak University of Technology,  
Namestie Slobody 17, 812 31 Bratislava,  
Slovak Republic*

## 1. Introduction

Sintering of powders is one of the most important processes for the development of polycrystalline materials. The microstructure of a material is of fundamental importance in the processing of ceramics and metals since it affects the physical properties of the final product. Progress in our ability to satisfactorily predict microstructure and its properties has been quite slow owing to complexity of physical processes involved. The complete prediction of microstructural development in polycrystalline solids as a function of time and temperature is a major objective in materials science.

Grain size is a very important characteristic for evaluating properties of the materials, especially when we need to balance different ones [1]. During the sintering of polycrystalline materials the normal grain growth obeys the basic law

$$R = k \cdot t^n, \quad (1)$$

where  $R$  is an average grain size,  $k$  is a constant with Arrhenius temperature dependence,  $t$  is sintering time and  $n$  is a kinetic grain growth exponent. However the grain growth is influenced by many other input parameters.

Recently, computer simulation techniques have been developed, which can successfully incorporate many aspects of the grain interactions and can predict the main features of the microstructure [2-10]. The aim of simulation of polycrystalline grain growth is to approximate to the highest degree to the real structures. Relations between Monte Carlo simulations and real structures have been studied in [11]. A procedure for the simulation and reconstruction of real structures in crystalline solids has been presented in [12]. Experimental and computational studies of grain growth for other various types of materials have been carried out, e.g. in [13-14].

The most realistic correspondence between the evolution of real and simulated structure was achieved by Monte Carlo simulations. Monte Carlo simulation is a stochastic Markov process that generates a sequence of configurations of lattice site states. Trial states are generated from a random distribution and are either accepted or rejected with a probability given by the Boltzmann factor.

The generalized  $Q$ -state Potts spin model is applied to the simulation procedure. The structure development is mapped onto the two-dimensional or three-dimensional discrete simulation lattice. An area element of microstructure is represented by one lattice site and is assigned a random number  $Q_i$  ( $1 < Q_i < Q$ ) called orientation or spin. Grain boundary lies between two adjacent sites with different orientation. The energy of a lattice site is given by the Hamiltonian

$$E = J \sum_{j=1}^n (1 - \delta_{Q_i, Q_j}), \quad (2)$$

where  $J$  is a positive constant,  $Q_i$  is the orientation of the  $i$ -th lattice site,  $Q_j$  is the orientation of the  $j$ -th neighboring lattice site,  $\delta_{Q_i, Q_j}$  is the Kronecker delta. The sum is given over  $n$  vicinal lattice sites.

During the simulation procedure the  $i$ -th lattice site orientation is generated randomly and its energy  $E_1$  is calculated according to (2). Then a new random orientation is given to the  $i$ -th lattice site and energy  $E_2$  after reorientation is again calculated. The reorientation is accepted when  $E_2 < E_1$ . Otherwise the reorientation is accepted with the probability

$$P \approx \exp\{-\Delta E/kT\}, \quad (3)$$

where

$$\Delta E = E_2 - E_1, \quad (4)$$

$k$  is the Boltzmann constant and  $T$  is the temperature. The term  $J/kT$  can be replaced by  $\alpha$  also called temperature factor and for the final probability of the reorientation acceptance one obtains

$$P \approx \exp\{-\alpha d\}. \quad (5)$$

If the 2D lattice consists of  $N \times N$  lattice sites,  $N \times N$  reorientation attempts represent a time unit called Monte Carlo step (MCS). On the other hand for 3D simulation array  $N \times N \times N$  reorientation attempts represent one MCS. In all simulation types described in the contribution the lattice sites can be arranged either in square or hexagonal configuration. The type of the simulation lattice is one of the input parameters before the simulation starts. The influence of this parameter on simulated structure and average grain size was studied in [15].

As mentioned above the initialization of the simulation lattice can be based on random number orientations. However, instead of random number one can employ also experimental orientation. Then the input microstructure can be an experimental one measured either by EBSD (Electron Back Scattered Diffraction) [16] to simulate grain growth or by TEM (Transmission Electron Microscope) to simulate primary recrystallization [17]. Then because the grain orientation is known the grain boundary nature is also known and then its energy can be adjusted (see e.g. [18]). The simulation procedure is universal and the initial simulation lattice can be obtained also from other devices e.g. from REM (Reflection Electron Microscope) [15].

## 2. Normal grain growth simulations

### 2.1 Monophase grain growth

Generally, the simulation algorithm of grain growth is based on the tendency of lattice points to achieve minimum energy. This elementary algorithm of monophase structure development was described in detail, e.g. in [15], [19-20]. An example of grain growth simulation on the 3D square simulation lattice with input parameters  $N = 100$ ,  $Q = 50$ ,  $\alpha = 5$ ,  $t = 1000$  MCS is shown in various display modes in Fig 1.

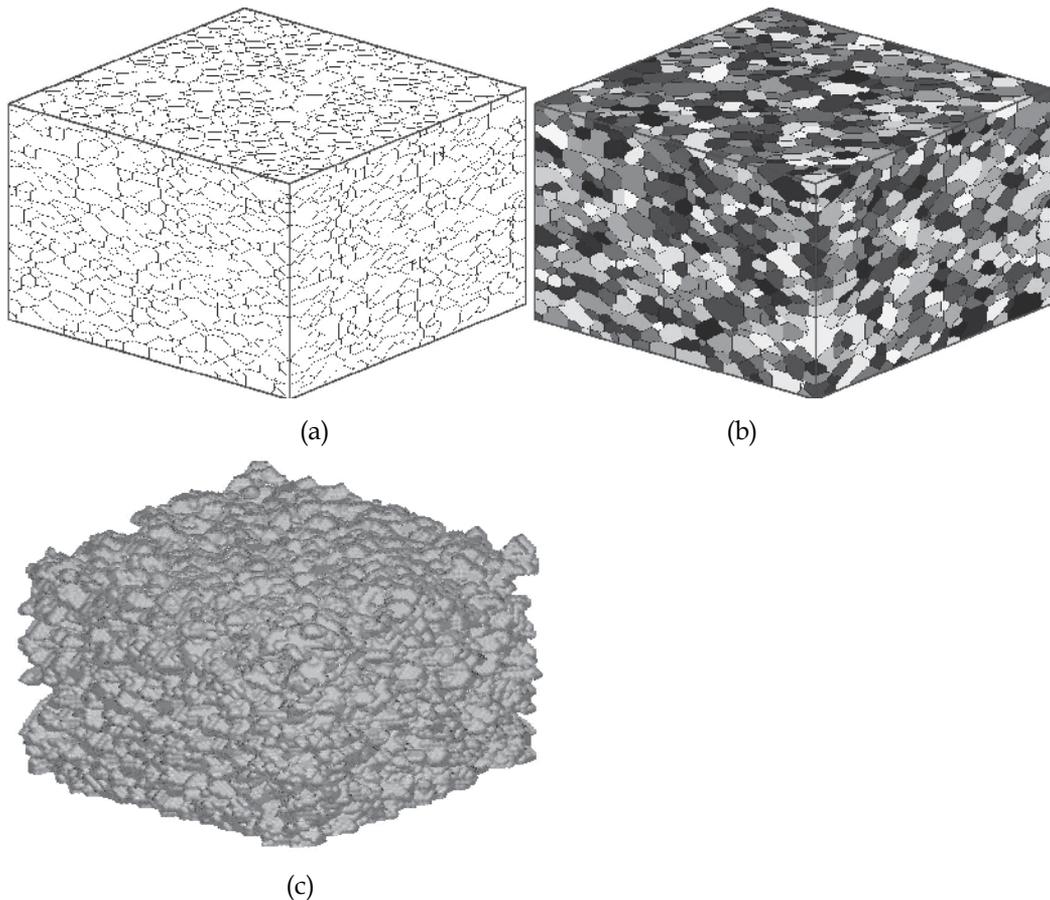


Fig. 1. Grain growth simulated on the 3D square simulation lattice with input parameters  $N = 100$ ,  $Q = 50$ ,  $\alpha = 5$ ,  $t = 1000$  MCS shown in simple display mode a), in shaded grains mode according to  $Q$  b), and shaded surface mode c).

### 2.2 Grain growth with presence of static second phase

The static second phase do not participate in the energy interaction. If during the simulation the lattice point with the orientation  $Q_s$  is randomly chosen this trial is ignored. The simulation continues with another trial. Consequently the positions of the static second phase lattice sites before and after simulation procedure are the same [22-31]. The static second phase lattice points can be arranged either in the form of grain inclusions, whiskers,

fibers. The influence of the input parameters on the simulated microstructure development in Monte Carlo simulations for both monophasic materials and materials containing static second-phase particles has been studied in [32]. An example of 3D grain growth simulation with the static second phase in the form of grains (5%) *a*., in the form of whiskers (5%) *b*., and in the form of fibers (10%) *c*. is given in Fig. 2.

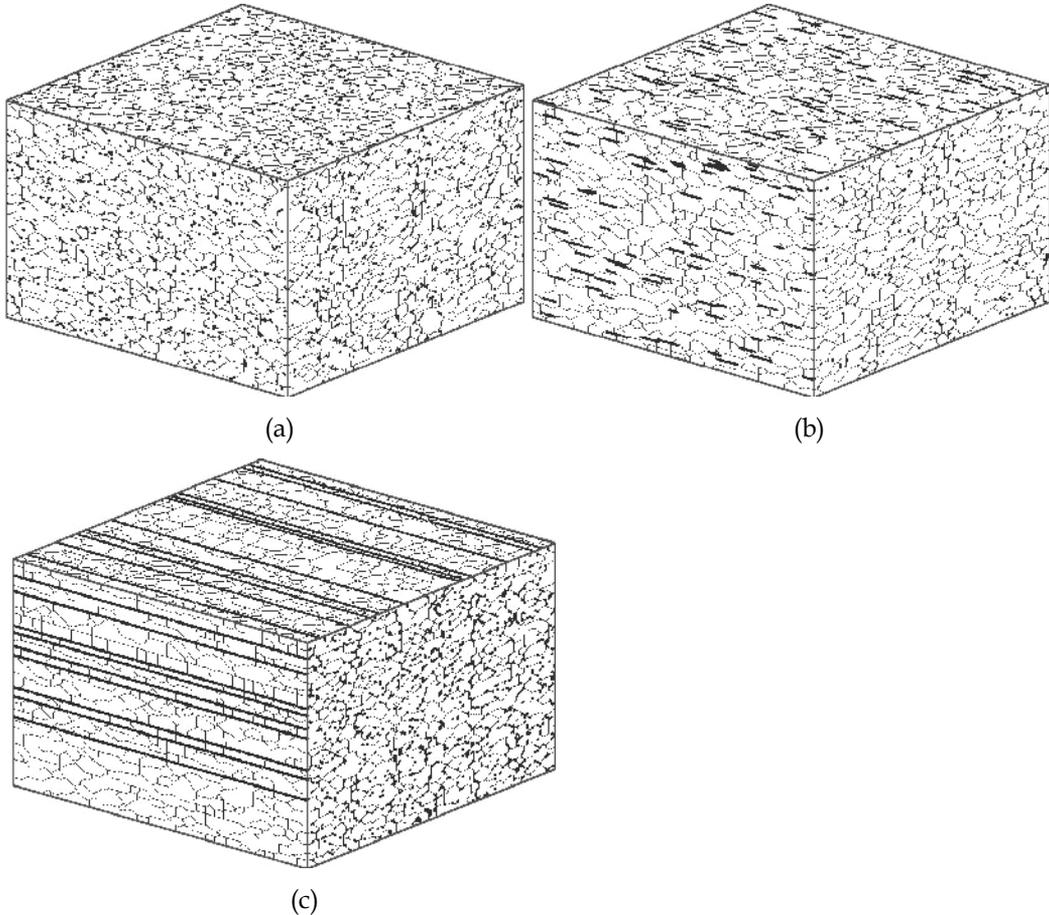


Fig. 2. Grain growth simulated on the 3D simulation lattice with input parameters  $N = 100$ ,  $Q = 50$ ,  $\alpha = 5$ ,  $t = 1000$  MCS with the static second phase in the form of grains (5%) *a*), in the form of whiskers (5%) *b*), and in the form of fibers (10%) *c*).

### 2.3 Grain growth in two-phase materials

When simulating grain growth in two-phase materials two types of grains with two different melting temperatures should be taken into account [28-29]. These parameters are represented by two temperature coefficients  $\alpha$  ( $\alpha = J/kT$ ), one for each phase. Then the simulation is carried out analogously to that described, e.g. in [15], [19] with different  $\alpha$  for each phase. An example of biphasic grain growth simulation is illustrated in Fig. 3. Due to both, smaller volume of the second phase grains and smaller  $\alpha$  the grains of the second phase are smaller.

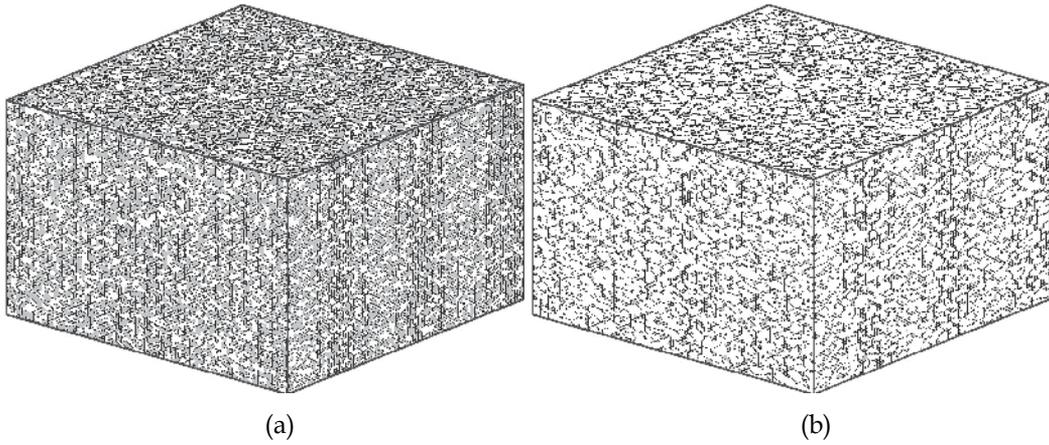


Fig. 3. Biphasic grain growth simulated on the 3D simulation lattice with input parameters  $N = 100, Q = 50, t = 1000$  MCS with the second phase volume = 50%,  $\alpha_1 = \alpha_2 = 5$  a), and with the second phase volume = 20%,  $\alpha_1 = 5, \alpha_2 = 1$  b).

#### 2.4 Grain growth with presence of liquid phase

There are many materials, which are prepared by the sintering process under the existence of a liquid phase [30]. In what follows the computer simulation algorithm of the grain growth in the presence of liquid phase is proposed:

- the required percentage of lattice points belonging to the solid phase is initialized randomly with the orientations from the interval  $\langle 1, Q \rangle$ . The rest of lattice points belonging to the liquid phase are initialized with the orientation  $Q_L$ ;
- if the chosen lattice point belongs to the solid phase the reorientation trial follows the algorithm given in [15];
- if the chosen lattice point belongs to the liquid phase with coordinates  $(i_1, j_1)$  so called “mass transfer algorithm” is applied :
  - a. using “random walking algorithm without back step” algorithm [30] we find the first point of the solid phase with coordinates  $(i_2, j_2)$  and orientation  $Q_{Sol}$ ;
  - b. the energy balance at the liquid phase point  $(i_1, j_1)$  is calculated  $-E_{A1}$ ;
  - c. the energy balance at the solid phase point  $(i_2, j_2)$  is calculated  $-E_{A2}$ ;
  - d.  $E_A = E_{A1} + E_{A2}$ ;
  - e. temporarily the solid phase point  $(i_2, j_2)$  is replaced by liquid point and energy balance  $E_{B2}$  is calculated;
  - f. successively for  $k \in \langle 1, Q \rangle$  we calculate the energy balance at the point  $(i_1, j_1)$  and find the smallest  $E_{B1opt}(k)$ ;
  - g.  $E_B = E_{B1opt}(k) + E_{B2}$ ;
  - h. if  $E_B < E_A$  the exchange is accepted. Otherwise the old orientations are left unchanged.

For illustration we introduce the structure development in the presence of liquid phase (shaded lattice points) that was simulated for  $N = 200, Q = 50, t = 100$  MCS,  $\alpha = 5, \gamma_{SL} = 50, \gamma_{SS} = 50$  with 10 % (Fig. 4a) and 40 % (Fig. 4b) of liquid phase  $L$ , respectively. In Fig. 5 we present an example of 3D simulation with liquid phase.

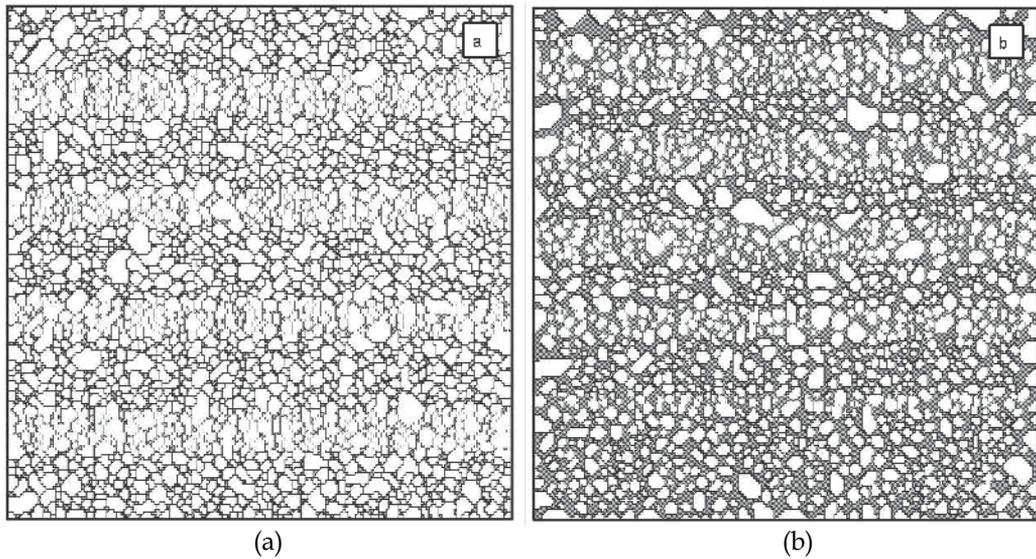


Fig. 4. Grain growth in the presence of liquid phase simulated on the square simulation lattice with input parameters  $N = 200, Q = 50, \alpha = 5, t = 100$  MCS,  $\gamma_{SL} = 50, \gamma_{SS} = 50, L = 10\%$  a) and  $L = 40\%$ , b).

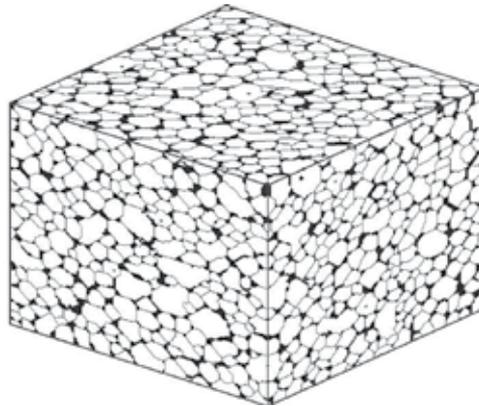


Fig. 5. 3D grain growth simulation in the presence of liquid phase with input parameters  $N = 100, Q = 50, \alpha = 5, t = 200$  MCS,  $L = 20\%, \gamma_{SL} = 50, \gamma_{SS} = 50$ .

### 2.5 Grain growth in the presence of gaseous phase

During the simulation of the structure development of the materials with the presence of dynamical pores, we considered simultaneously the energy balance point of view of solid particles sites as well as the direction of the pores motion aspect [31], [33-35]. In other words along with the simulation of the grain growth through the use of the above given procedures we have to simulate the migration of pores as well. The algorithm of the pore migration involves

- the determination of the direction of the motion
- the calculation of eventual change of the pore position in this direction.

**2.5.1 Energy balance calculation during pore migration**

The kinetics of the pores is realized via the exchange of the orientation of the lattice point *A* by the orientation of some of neighboring points, e.g. by the orientation of the point *B*. Using (2) we calculate the energy of the pore site *A* -  $E_{1A}$  and the energy of the site *B* -  $E_{1B}$ . Then

$$E_1 = E_{1A} + E_{1B}; \tag{6}$$

- we exchange points *A* and *B*;
- again using (2) we calculate the energies of both exchanged points -  $E_{2A}, E_{2B}$ . Then

$$E_2 = E_{2A} + E_{2B}; \tag{7}$$

- the difference of the energies before and after the exchange of the points *A* and *B* is

$$\Delta E = E_2 - E_1; \tag{8}$$

- if  $\Delta E \leq 0$ , the exchange of the sites *A* and *B* is accepted with the probability equal to 1, otherwise it is accepted with the probability

$$P \approx \exp\{-\beta\Delta E\}, \tag{9}$$

where  $\beta$  is temperature coefficient of the pore motion.

**2.5.2 Direction of the pore motion**

We have studied four models of the pore migrations using different approaches determining the direction of the pore motion. In all the algorithms let us assume that during the simulation we have randomly chosen lattice site *A* with  $Q_A = Q_P$ .

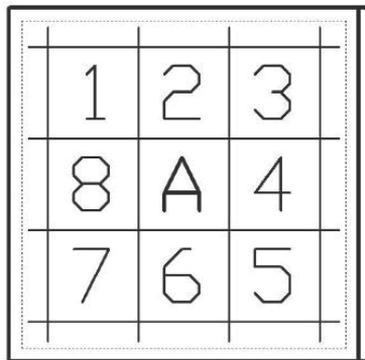


Fig. 6. A part of square simulation lattice with pore lattice site surrounded by lattice points denoted 1 ÷ 8 . Double line denotes the nearest edge of the simulation lattice.

**2.5.2.1 Stochastic model of the pore motion**

In this model, the motion of pores is allowed with equal probability in all directions. The algorithm of the pore motion simulation is as follows:

- in the first step let us denote the 8 lattice points neighboring with the chosen pore site *A* by numbers from 1 to 8 according to Fig. 6. Let us assume that the right side of the simulation array (denoted by double line) is the nearest edge (from all 4 edges of the array) to the site *A*.

- let us generate the random number (uniform distribution) from the interval  $\langle 1,8 \rangle$  determining the point  $B$  and thus the direction of the eventual pore motion;
- then the energy balance calculation is carried out between these two points according to the algorithm presented in the section 2.5.1. In this model, the motion of pores is allowed with equal probability in all directions. The algorithm of the pore motion simulation is as follows:

**2.5.2.2 Probability model of the pore motion**

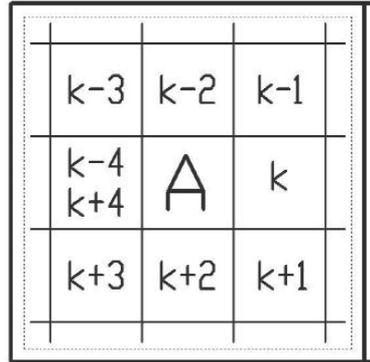


Fig. 7. Distribution of the lattice sites neighboring with the pore site  $A$  and denoted  $k-4 \div k+4$  in the lattice space, where  $k$  is the direction to the nearest simulation lattice edge (double line).

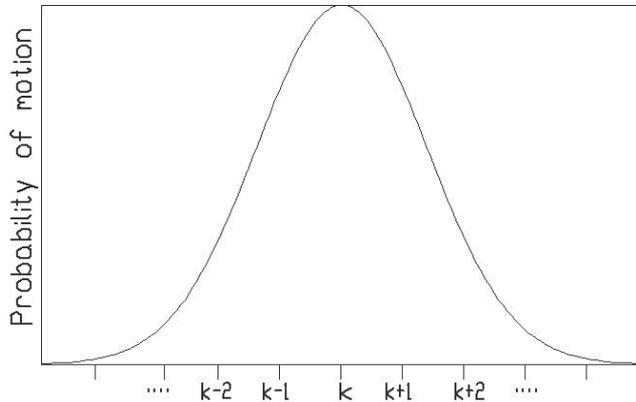


Fig. 8. Distribution of the lattice sites neighboring with the pore site  $A$  and denoted  $k-4 \div k+4$  according to the Gaussian distribution.

In this model, the probability of the pore motion is determined by Gaussian distribution around the direction to the nearest edge of the simulation lattice. The algorithm of the pore motion simulation is as follows:

- we determine the nearest edge of the simulation lattice. The smallest distance to an edge of the lattice is

$$c = \min(X_A, N - X_A, Y_A, N - Y_A), \tag{10}$$

where  $X_A, Y_A$  are the coordinates of the point  $A$  (see Fig. 7). The nearest edge is denoted by double line on the right side of the simulation lattice. We select the direction satisfying (10) and we denote it  $k$ ;

- other lattice points neighboring to the site  $A$  are denoted according to Fig. 7;
- Gaussian random number generator (with  $\sigma$  as input parameter) generates random number from the interval  $\langle k-4, k+4 \rangle$  according to Fig. 8;
- based on this number and using the notation from the Fig. 7 we select neighboring point  $B$ ;
- then the energy balance calculation is carried out between these two points according to the algorithm presented in the section 2.5.1.

### 2.5.2.3 Motion in directions $\langle k-2, k+2 \rangle$ with equal probability - edge model

Using (10) we determine the direction  $k$ . We shall suppose that the pore point  $A$  can interact only with one of the five possible lattice sites denoted in Fig. 9 as  $k-2, k-1, k, k+1, k+2$ . They are symmetrically distributed around the basic direction given by position of the site  $k$ .

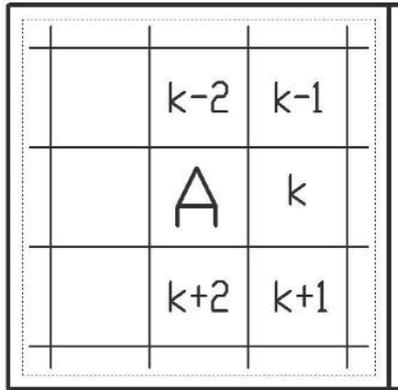


Fig. 9. Lattice point  $A$  surrounded by points denoted  $k-2 \div k+2$ . To define interacting point  $B$  we generate a random number from the interval  $\langle 1, 5 \rangle$  that corresponds to the sites  $k-2, k-1, k, k+1, k+2$ . Then the energy balance calculation is carried out with this point.

### 2.5.3 Results of simulations of pore migration

In Fig. 10a we present final structure after grain growth simulation along with pore migration according to the stochastic model. Due to uniform distribution of the pore motion in all directions, relatively large clusters of pores were enclosed inside of the material. Moreover large amount of small, one point pores (pores of the first generation), remained in the material as well.

In the probability model, it is possible to control the Gaussian distribution of the pores motion. An example of the simulation employing this model for  $\sigma=2$  is given in Fig. 10b. Only few clusters of pores remained encapsulated. They have regular elliptical shape. One can notice the bent square of solid material in the simulation lattice.

Finally in Fig. 11a we show the resulting structure with pores motion simulation according to the edge model (after 1000 MCS,  $\beta=1000$ ). The majority of pores left the structure and moved to the edges of the simulation lattice. Fewer clusters remained encapsulated inside of the solid material than in the stochastic model. This model like probability model allows to simulate shrinking of pores along with their motion to the edge of the lattice. We can go on with the simulations and change the temperature coefficient of pore migration to  $\beta = 2$ . The structures after 1020 MCS, 1040 MCS and after 5000 MCS are shown in the Figs. 11b, 11c, and 11d respectively. One can see that the encapsulated pores disappeared from the material and the square lattice was straightened.

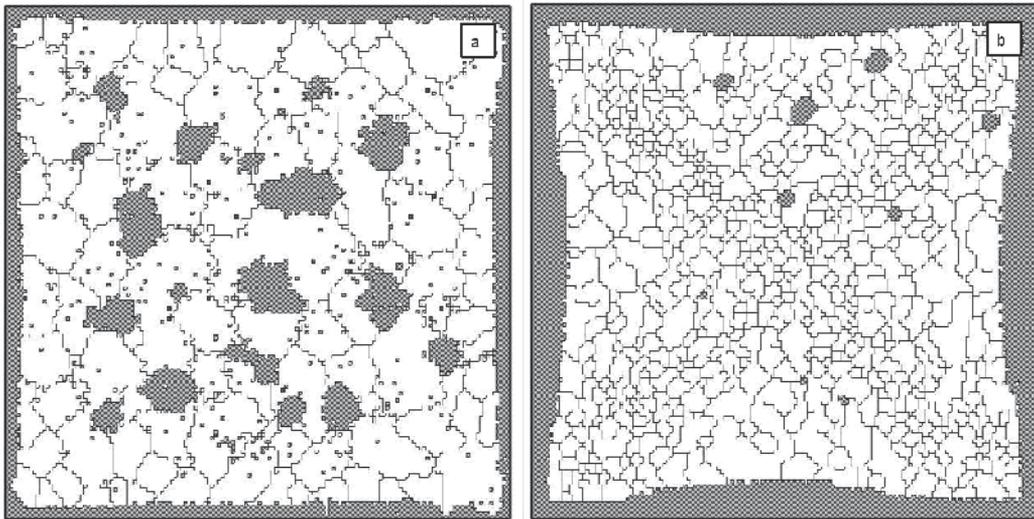


Fig. 10. Grain growth with the mobile pores simulated on the square simulation lattice according to the stochastic model a), the probability model ( $\sigma = 2$ ) b) with input parameters  $N = 150, Q = 40, \alpha = 1000, \beta = 1000, P = 20\%, t = 5000$  MCS.

One can ask why the bent square of the solid material in Fig. 10b is greater than in Fig. 10a and why it completely disappears in Fig. 11d. The difference between both models consists in different probabilities of pores motion. Consequently every model has different speed of the pores motion to the edge of simulation lattice. The aim was to carry out several simulations for both models and to find out which model corresponds to real structures. Actually the sintered ceramic pellets are in fact bent in a way the proposed simulation models indicate.

In Figs. 11 a, b, and c we decreased the temperature coefficient to  $\beta = 2$ . Due to this the pores tend to move to their closest edges of the simulation lattice. In Figs. 11b and c we present intermediate results after 1020 and 1040 MCS, respectively. When we increase dramatically the simulation time to 5000 MCS all pores leave the solid material. However the simulation process of monophasic grain growth in solid material goes on. Due to the finite simulation lattice the pores cannot move in the perpendicular direction towards the edges. They are forced to move along the edges and as a consequence the square lattice is straightened.

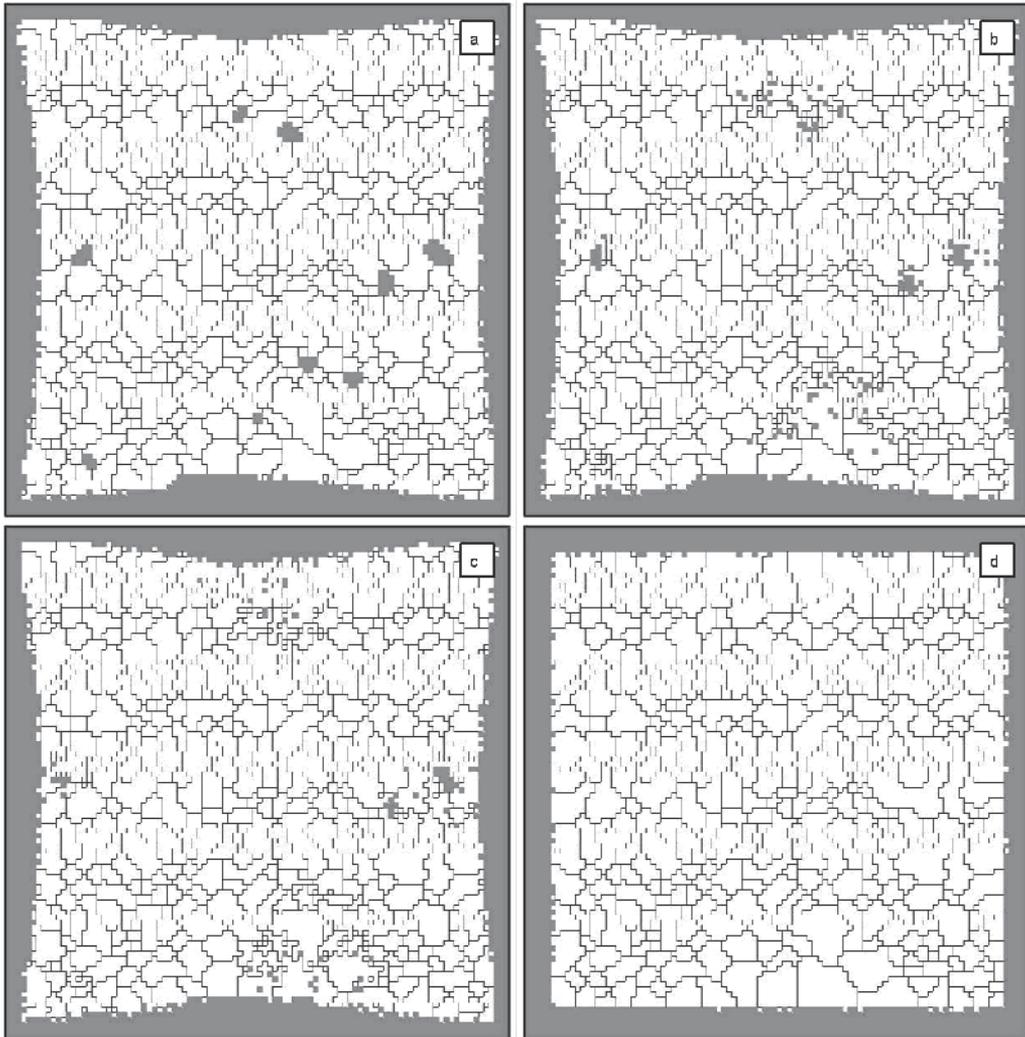


Fig. 11. Grain growth with the mobile pores simulated on the square simulation lattice according to the edge model with input parameters  $N = 100$ ,  $Q = 60$ ,  $\alpha = 1000$ ,  $\beta = 1000$ ,  $P = 20\%$ ,  $t = 1000$  MCS a), then  $\beta = 2$ ,  $t = 1020$  MCS b),  $t = 1040$  MCS c) and  $t = 5000$  MCS.

### 3. Oriented grain growth simulations

#### 3.1 Oriented grain growth in one direction

During the simulation the excess of energy in preferred direction, which determines the grain boundary curvature, can be influenced by changing the value  $J$  in the Hamiltonian (2) in dependence of the neighboring sites [36-38]. It means that neighboring sites contribute with different weights to the Hamiltonian in (2). Hence the Hamiltonian for oriented structures can be written as

$$E = -\sum J(\delta_{Q_i Q_j} - 1), \quad (11)$$

where

$$J = \sum_{j=1}^N J_j . \quad (12)$$

The value  $J_{Pr}$  of the lattice points in the preferred direction equals to the multiple of  $J_{N_i}$  in the non-preferred direction ( $N_i \neq Pr$ ). In practice the preferential grain growth is given by the weights

$$W_i = \frac{J_P}{J_{N_i}} . \quad (13)$$

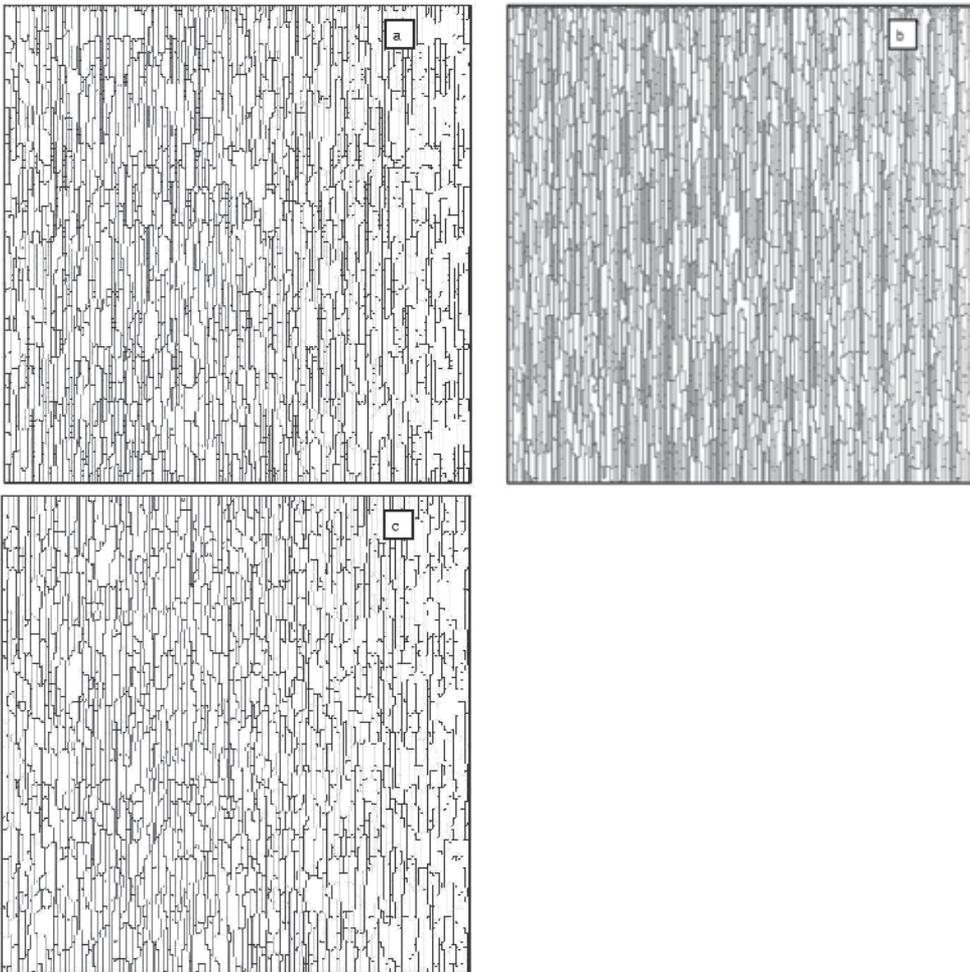


Fig. 12. Oriented grain growth simulated on the square simulation lattice with input parameters  $N = 200$ ,  $Q = 50$ ,  $\alpha = 5$ ,  $t = 1000$  MCS using square model a), cross model b) and elliptical model c). Direction of preferred growth is  $y$ .

In [15] for the square lattice three various algorithms to specify preferred direction were proposed:

1. *two-weights square model* - the weight of grain growth in preferred direction is  $W_1$ , (horizontal or vertical) the weights of other lattice points neighboring with the point being evaluated are  $W_2$ .
2. *two-weights cross model* - it allows the evaluated site to interact only with four neighboring sites in horizontal and vertical directions - two points in the preferred direction have the weights  $W_1$  and two points in the other allowed positions have the weights  $W_2$ . The neighbors in diagonal directions do not participate in the energy interaction, i.e., their weights are equal to zero.
3. *three-weights elliptical model* - in this model we have proposed three directions - horizontal, vertical and diagonal. The weight  $W_3$  in diagonal directions is defined by ellipse with semi-axes  $W_1$  and  $W_2$  as

$$W_3 = W_1 W_2 \sqrt{2 / (W_1^2 + W_2^2)} .$$

To illustrate the influence of the model on the shape of grains in Figs. 12 a - c we present the results of the oriented grain growth with preferred direction  $y$  simulated with the square (a), cross (b) and elliptical (c) models, respectively. In Fig. 13 we present oriented grain growth simulated on 3D simulation lattice using elliptical model with preferred direction  $z$  (a), and preferred directions  $y, z$  (b).

### 3.2 Anisotropic grain growth

#### 3.2.1 Anisotropic grain growth in solid state

While in the oriented grain growth the preferred direction of the growth is the same for all grains in case of anisotropic structures it is related only to a restricted number of grains [39]. The geometrical anisotropic grain growth can be due to crystallographic effects [40]. In the simulation procedure the direction of growth of an anisotropic grain is random. For each anisotropic grain, we assign an arbitrary direction of the growth. For square simulation lattice it is one of the four directions and for triangular simulation lattice it is one of the three directions. Then we proceed according to the following algorithm:

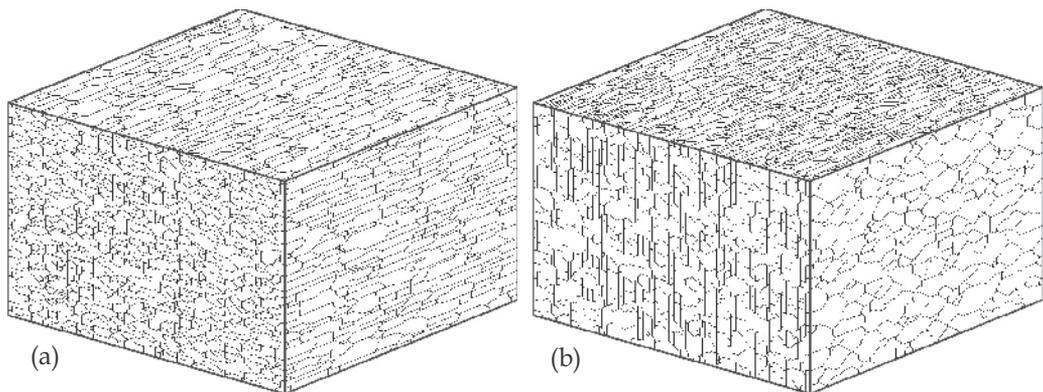


Fig. 13. Oriented grain growth simulated on 3D simulation lattice with input parameters  $N = 100, Q = 50, \alpha = 5, t = 1000$  MCS using elliptical model with weights 1:1:10 a), and 1:10:10 b).

- orientation  $Q$  is divided into two intervals  $\langle 1, Q_E \rangle$  and  $\langle Q_E + 1, Q \rangle$  proportionally to desired percentage  $p_E$  of anisotropic grains, i.e.,  $Q_E = Q - p_E \cdot Q / 100$ ;
- anisotropic lattice points are randomly assigned orientations from the interval  $\langle 1, Q_E \rangle$ ;
- the rest of lattice points, obeying normal grain growth law, are randomly assigned orientations from the interval  $\langle Q_E + 1, Q \rangle$ ;
- for lattice points belonging to normal grains we apply the algorithm described in [15];
- for lattice points belonging to anisotropic grains, we apply the algorithm described in section 3.1 with preferred grain growth direction appertaining to the given orientation of the anisotropic grain.

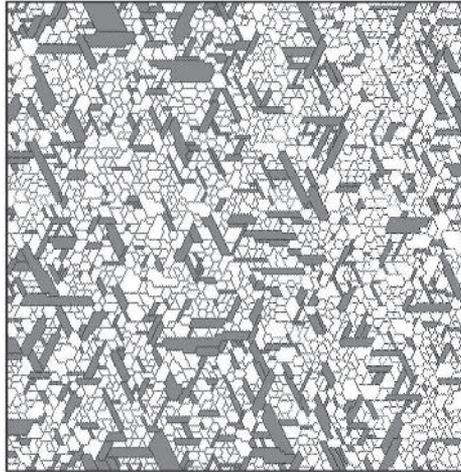


Fig. 14. Anisotropic grain growth according to the elliptical model simulated on the hexagonal simulation lattice with input parameters  $N = 150, Q = 50, \alpha = 5, A = 10\%$ ,  $t = 1000$  MCS and  $W1 : W2 : W3 = 1 : 1 : 20$ .

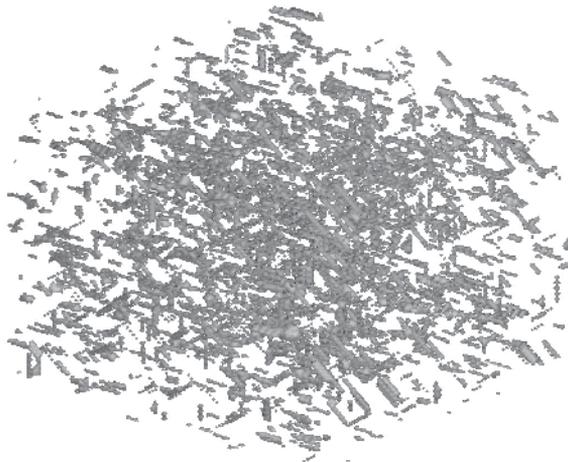


Fig. 15. 3D anisotropic grain growth according to the elliptical model with input parameters  $N = 100, Q = 250, \alpha = 50, A = 5\%$ ,  $t = 1000$  MCS and  $W1 : W2 : W3 = 30 : 1 : 1$ .

In Fig. 14 we show anisotropic grain growth, which was simulated on the hexagonal lattice. The elliptical simulation model with weights ratios  $W1 : W2 : W3 = 1 : 1 : 20$  and with 10% of anisotropic grains (A) (shaded lattice sites) has been chosen. Similar example of 3D anisotropic simulation is given in Fig. 15.

**3.2.2 Anisotropic grain growth in liquid phase**

The above presented simulation algorithm of the grain growth in the presence of liquid phase is dealing with the growth behavior under isotropic energy of solid/liquid interface  $\gamma_{SL}$ . However, in polycrystalline materials there exist material systems (ceramics, cermets, tungsten carbide,  $\alpha$  - alumina, etc), which have the anisotropic behavior of particles during liquid phase sintering [30], [41-42]. If the neighbor of a solid particle is the simulation site corresponding to the liquid phase the energy balance is calculated according to the following algorithm:

- for energies of the interface between solid particles and a liquid phase  $\gamma_{SL}$  ( $\gamma_{SL} \in (0,1)$ ) and between solid and solid particles  $\gamma_{SS}$  ( $\gamma_{SS} \in (0,1)$ ) it holds

$$\gamma_{SS} > \gamma_{SL};$$

- let us denote

$$a = \gamma_{SL}$$

$$b = (\gamma_{SS} - \gamma_{SL})/3;$$

- the direction of the interaction for square lattice

$$direction = Q \bmod 4$$

and for hexagonal lattice

$$direction = Q \bmod 3,$$

where Q is the orientation of the solid particle lattice site. The  $direction_i$  is chosen according to the position of the neighbor and the chart shown in Fig. 16a, e.g. for the point B in Fig. 16b the  $direction_i = 2$ .

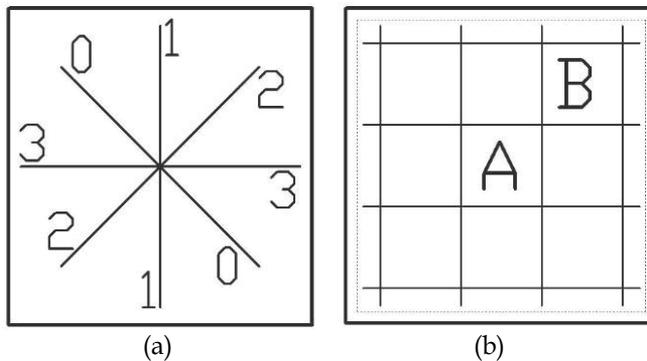


Fig. 16. The chart of possible positions of the neighbors a) and corresponding point B if the position was chosen 2 b).

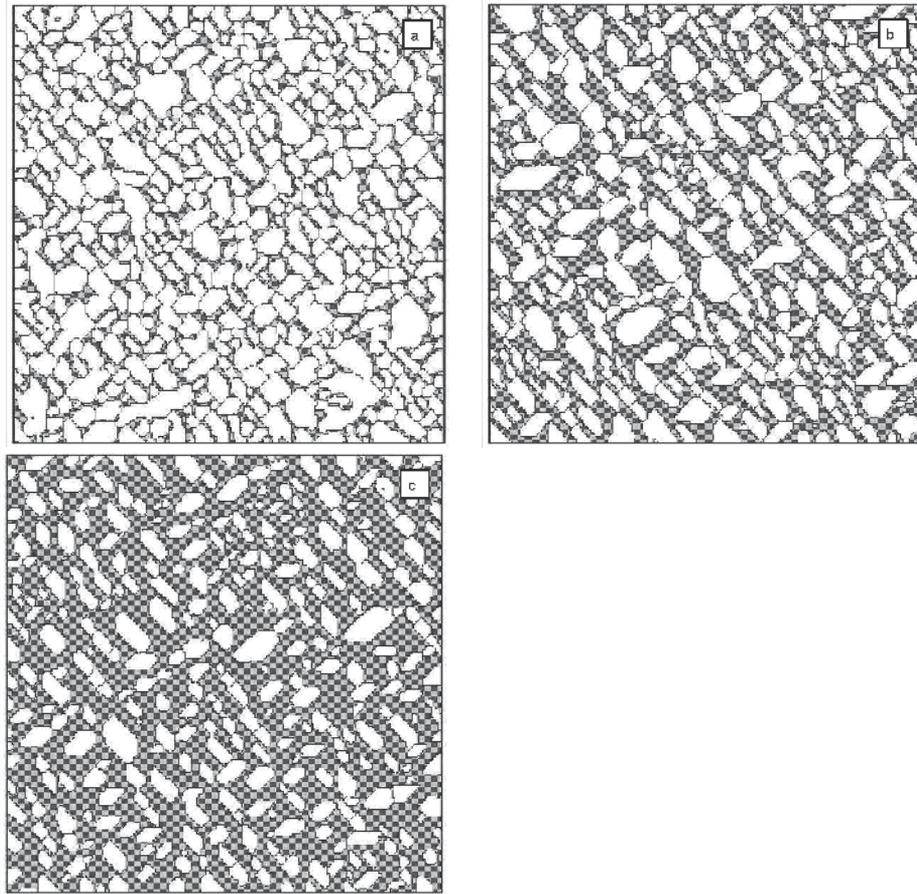


Fig. 17. Anisotropic grain growth in the presence of liquid phase simulated on the hexagonal simulation lattice with input parameters  $N = 200, Q = 50, \alpha = 5, A = 100\%, t = 1000$  MCS,  $\gamma_{SL} = 10, \gamma_{SS} = 90$  and  $L = 20\%$  a), 40 % b), 60 % c).

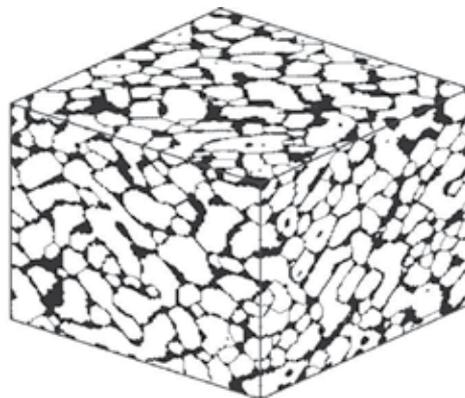


Fig. 18. 3D grain growth simulation in the presence of liquid phase with input parameters  $N = 100, Q = 50, \alpha = 5, t = 200$  MCS,  $L = 20\%, \gamma_{SL} = 10, \gamma_{SS} = 90$ .

- then the increase of the energy in energy balance calculation around the solid particle for the  $direction_i$  is

$$a + b * (3 - |direction - direction_i|) .$$

The simulated structures with anisotropic grain growth in liquid phase are shown in Fig. 17 with concentrations of liquid  $L = 20\%$  a),  $40\%$  b) and  $60\%$  c, respectively. In Fig. 18 we give an example of 3D anisotropic grain growth simulation in the presence of liquid phase with input parameters  $N = 100, Q = 50, \alpha = 5, t = 200$  MCS,  $L = 20\%, \gamma_{SL} = 10, \gamma_{SS} = 90$ .

#### 4. Conclusion

In the contribution, we have given an account of various simulation algorithms of the grain growth in polycrystalline materials. We have presented sophisticated algorithms of pore migration, simulation of grain growth in presence of liquid, in oriented and anisotropic structures. All the algorithms have been extended to the three-dimensional simulation arrays. Different input parameters can influence the average grain size, which is very important parameter because it is closely connected with many properties of simulated structures. It can be obtained by scanning the whole simulation lattice using intercept length method. Histograms of the studied parameters are automatically recorded during the simulation.

The average grain size decreases with increasing number of orientations  $Q$ . It is a factor that refers to the particle size distribution in real powders. The simulation carried out for small value of  $Q$  results in small number of irregular grains. On the contrary, the high value of  $Q$  gives small and regular grains similar to the monodisperse particle distribution. Another important parameter that influences the average grain size is simulation time. The study of dependence of this parameter on time has shown that to obtain stable simulation structure the simulation time 1000 MCS is sufficiently long. In [32] the detailed study of the dependence of the size of simulation lattice, type of simulation lattice (square or hexagonal), number of orientations  $Q$ , temperature coefficient  $\alpha$ , etc. on the average grain size was carried out and the results discussed.

The simulation algorithms presented above were implemented in the software package WinSimul, which was developed at the Institute of Physics, Slovak Academy of Sciences. It allows to simulate the structure development, to evaluate the simulated structures, to display lattice during simulation or to record display frames in time in the form of AVI files. It also can display simulation results in the form of average grain size, average area and neighbors (topological) histograms or time dependences of these parameters. It is possible to display several simulations or simulation results simultaneously for various input parameters.

A modular structure of the program WinSimul provides a great flexibility of simulation configurations. Presented work shows only some possible combinations from many others, which may occur in practice.

#### 5. References

- [1] Ling, S. and Anderson, M. P.: Jom, 44, 1992, p.30.
- [2] Matsubara, H. – Furukawa, K. – Brook, R. J.: In: Proceedings 4<sup>th</sup> Euro-Ceramics Society Conference, Eds.: B.S.Tranchina, A.Bellosi, Gruppo Editoriale Faenza Editrice, Riccione, Italy, 1995, p. 597.

- [3] Kurtz, S. K. – Carpay, F. M. A.: *J. Appl. Phys.* 51, 1980, p.5745.
- [4] Anderson, M. P. – Grest, G. S. – Srolovitz, D. J.: *Phil. Mag. B* 59, 1989, p. 293.
- [5] Rollet, A. D. – Luton, M. J. – Srolovitz, D. J.: *Acta Metall. Mater.* 40, 1992, p.43.
- [6] Petzak, P. – Luton, M. J.: *Springer Proceedings in Physics*, 76, 1993, p. 46.
- [7] Hassold, G. N. – Holm, E. A.: *Computers in Physics*, 7, 1993, p. 97.
- [8] Morháčová, E.: *Cryst. Res. Technol.*, 29, 1994, K99.
- [9] Blikstein, P. – Tschiptschin, A. P.: *Mat. Res.* 2, 1999, p. 133.
- [10] Miller, R. S. – Cao, G. – Grujicic, M.: *J. Mater. Synth. Process*, 9, 2001, p. 329.
- [11] Morháčová, E.: *Cryst. Res. Technol.*, 30, 1995, K9.
- [12] Rahman, S. H.: *Acta Cryst.* 49, 1993, p. 56.
- [13] Blikstein, P. – Tschiptschin, A.P.: *Mat. Res.* 2, 1999, p. 133.
- [14] Tajika, M. – Nomura, H. – Matsubara, H. – Rafaniello, W.: *J. Ceram. Soc. Japan* 109, 2001, p. 288.
- [15] Morháč, M. – Morháčová, E.: *Cryst. Res. Technol.*, 35, 2000, p. 117.
- [16] Baudin, T. – Paillard, P. – Penelle, R.: *Script. Mater.* 36, 1997, p. 789.
- [17] Baudin, T. – Julliard, F. – Paillard, P. – Penelle, R.: *Script. Mater.* 43, 2000, p. 63.
- [18] Caleyó, F. – Baudin, T. – Paillard, P. – Penelle, R.: *Script. Mater.* 46, 2002, p. 829.
- [19] Tikare, V. – Cawley, J. D.: *J. Am. Ceram. Soc.* 81, 1998, p. 485.
- [20] Raabe, D. – Roters, F. – Barlat, F. – Chen, L. Q.: *Continuum Scale Simulation of Engineering Materials: Fundamentals – Microstructures – Process Applications*, Weinheim, Wiley-VCH Verlag GmbH & Co. KGaA, 2004
- [21] Yabushita, S. – Hatta, N. – Kikuchi, S. – Kokado, J.: *Script. Metall. Mater.* 19, 1985, p. 853.
- [22] Doherty, R. D. – Li, K. – Kashyup, K. – Rollet, A. D. – Anderson, M. P.: In: *Proc. of Tenth Risø International Symposium on Metallurgy and Material Science : Materials Architecture*, Eds.: J.B. Bilde-Sorensen et al., Risø National Laboratory, Roskilde, Denmark, 1989, p. 31.
- [23] Anderson, M. P. – Grest, G. S. – Doherty, R. D. – Li, K. – Srolovitz, D. J.: *Script. Metall. Mater.* 23, 1989, p. 753.
- [24] Hassold, G. N. – Holm, E. A. – Srolovitz, D. J.: *Script. Metall. Mater.* 24, 1990, p. 101.
- [25] Hazzeldine, P. M. – Oldershaw, R. D. J.: *Phil. Mag. A* 61, 1990, p. 579.
- [26] Olguín, A. et al.: *Phil. Mag. B* 81, 2001, p. 731.
- [27] Yoshikiro, S. – Yoshiyuki, S. – Hideiro, O.: *Script. Mater.* 55, 2006, p. 407.
- [28] Cahn, J.W. – Holm, E. A. – Srolovitz, D. J.: *Mat. Sci. Forum* 94-96, 1992, p. 141.
- [29] Zheng, Y. G. et al.: *Appl. Phys. Lett.* 88, 2006, 144103.
- [30] Matsubara, H. – Furukawa, K. – Brook, R. J.: In: *Proceedings 4<sup>th</sup> Euro-Ceramics Society Conference*, Eds.: B.S.Tranchina, A.Bellosi, Gruppo Editoriale Faenza Editrice, Riccione, Italy, 1995, p. 529.
- [31] Hassold, G. N. – Srolovitz, D. J.: *Scripta Metall. et Mater.* 32, 1995, p. 1541.
- [32] Morháčová, E. – Morháč, M.: *Kovove Mater.* 45, 2007, p. 105.
- [33] Chen, I. W. – Hassold, G. N. – Srolovitz, D. J.: *J. Amer. Ceram. Soc.* 73, 1990, p. 2865.
- [34] Morháčová, E.: *Cryst. Res. Technol.*, 30, 1995, K4.
- [35] Tikare, V. – Miodownik, M. A. – Holm, E. A.: *J. Am. Ceram. Soc.* 84, 2001, p. 1379.
- [36] Vaz, M. F. – Soares, A. – Fortes, M. A.: *Scripta Metall. et Mater.* 24, 1990, p. 2453.
- [37] Tavernier, P. – Szpunar, J. A. : *Acta Metall. et Mater.* 39, 1991, p. 549.
- [38] Tavernier, P. – Szpunar, J. A. : *Acta Metall. et Mater.* 39, 1991, p. 557.
- [39] Kunaver, U. – Kolar, D.: *Mat. Sci. Forum* 94-96, 1992, p. 785.
- [40] Baudin, T. – Paillard, P. – Penelle, R.: *Script. Mater.* 40, 1999, p. 1111.
- [41] German, R. M.: *Liquid phase sintering*, New York, Plenum Press, 1985.
- [42] Sebaugh, M.M. – Kerscht, I. H. – Messing, G. L.: *J. Am. Ceram. Soc.*, 80, 1997, p. 1181.

# Monte Carlo Simulations of Grain Growth in Metals

Sven K. Esche  
*Stevens Institute of Technology*  
USA

## 1. Introduction

The application of the Monte Carlo (MC) method to simulate the grain growth in metals originates from Potts' model for magnetic domain evolution (Potts, 1952), which generalized the two-state spin up or spin down ferromagnetic Ising model to systems with arbitrary spin degeneracy. Subsequently, the so-called n-fold method for expediting simulations of the time evolution of systems was developed (Bortz et al., 1975). Anderson and his co-workers were the first to introduce the Potts model into grain growth simulations, applying this method to model the grain growth kinetics (Anderson et al., 1984), grain size distribution and topology (Srolovitz et al., 1984a), influence of particle dispersions (Srolovitz et al., 1984b), anisotropic grain boundary energies (Grest et al., 1985) as well as abnormal grain growth (Srolovitz et al., 1985; Rollett et al., 1989; Rollett & Mullins, 1996). By incorporating specific elements corresponding to various microstructural processes into the basic algorithm, the MC method has been adapted to model for instance grain growth in two-phase materials (Holm et al., 1993) and composites (Miodownik et al., 2000), abnormal grain growth (Lee et al., 2000; Messina et al., 2001; Ivasishin et al., 2004), static recrystallization (Srolovitz et al., 1986; Srolovitz et al., 1988; Rollett et al., 1992a, Rollett & Raabe, 2001; Song & Rettenmayr, 2002)), dynamic recrystallization (Peczak, 1995; Rollett et al., 1992b) and sintering (Hassold, et al., 1990; Chen et al., 1990; Matsubara, 1999), and it has been demonstrated that such MC simulations are capable of reproducing the essential features of these microstructural phenomena. Nowadays, the MC method is often preferred to deterministic methods such as cellular automaton (Geiger et al., 2001) and phase-field models (Tikare et al., 1998) at the mesoscopic level, mainly due to its inherent simplicity and flexibility. More recently, the MC method has also been employed to predict the final microstructures in engineering applications (Yang et al., 2000; Yu & Esche, 2005).

For quite some time, numerous efforts geared toward improving the accuracy and efficiency of the conventional MC method have been reported in the literature (Radhakrishnan & Zacharia, 1995; Song & Liu, 1998, Yu & Esche, 2003a), aiming at providing the foundation for the application of the MC method in engineering practice. Various modifications of the conventional Monte Carlo (CMC) algorithm have been reported. For instance, an increase in processing speed of up to two orders of magnitude compared with the CMC algorithm were achieved in grain growth simulations by employing a modified MC algorithm (Yu & Esche, 2003a). Furthermore, this modified algorithm also led to an improved accuracy of the predicted grain growth exponent in the kinetic equations, particularly in small grain size

regimes (i.e., for grain sizes ranging from 1 to 15 lattice spacing units). This improvement of the accuracy of the simulation results also facilitated new insights into the performance of MC Potts models for simulating the microstructure evolution (Yu & Esche, 2003b, Yu & Esche, 2003c).

The kinetics of normal grain growth is governed by the following equation (Atkinson, 1998)

$$\langle R \rangle^m - \langle R \rangle_0^m = Mt \quad (1)$$

where  $\langle R \rangle$  is the average grain radius,  $\langle R \rangle_0$  is the initial average grain radius,  $m$  and  $M$  are constants and  $t$  is the time. If in the analyzed time interval the initial grain radius  $\langle R \rangle_0$  is negligible compared with  $\langle R \rangle$ , then the grain growth kinetics can be simplified to (Louat, 1974)

$$\langle R \rangle = Kt^n = M^n t^n \quad (2)$$

where  $n = 1/m$  is the grain growth exponent.

Despite the fact that there is ample theoretical (Burke & Turnbull, 1952; Mullins, 1956; Mullins & Vinals, 1989), computational (Weiare & Kermod, 1984) as well as experimental (Glazier et al., 1987) evidence that the kinetic law is parabolic (i.e.  $n = 0.5$ ), a substantial number of research results obtained from theoretical work (Rhines et al., 1974), computer simulations (Anderson et al., 1984) and physical grain growth experiments (Bolling & Winegard, 1958; Anderson et al., 1984) contradict the parabolic kinetic law. In physical experiments, the deviation from the parabolic kinetics can be explained by the varying grain boundary mobility and energy of the material samples or by the presence of a small number of second-phase particles in the material samples (Humphreys & Hatherly, 1996). However, the results of computer simulations with ideal (i.e., isotropic, single-phase) materials have also exhibited the deviation from the parabolic kinetic law. Even though theoretically the self-similarity of the microstructure evolution is a sufficient condition for the occurrence of the parabolic kinetic law in the curvature-driven grain growth process, some results showed that two-dimensional (2D) normal grain growth simulations of single-phase systems employing the MC Potts model result in a significantly lower exponent of  $n = 0.41$  in the small grain size regimes (Anderson et al., 1984; Grest et al., 1988), despite the fact that self-similarity was observed. The grain growth exponent was found to asymptotically approach  $n = 0.5$  at the later simulation stages when using large lattice systems (up to  $1000 \times 1000$  lattice points) with large final grain sizes (Grest et al., 1988) while in the small grain size regimes, it was believed that an unphysical finite-size effect is likely to dominate (Anderson et al., 1989). Furthermore, it was also demonstrated that normal grain growth is modeled almost precisely (i.e. with an exponent of  $n = 0.502$ ) if small clusters are excluded in calculations of the microstructure statistics (Ivasishin et al., 2003). The lower exponent  $n$  observed when including these small clusters was attributed to the presence of two competing length scales in the simulation, namely the mean grain size  $\langle R \rangle$  and the lattice spacing  $a_0$  (Kumar et al., 1987; Beenakker, 1988), and thus the effect of the finite lattice spacing and the self-similarity of the grain growth can be expected only in the limit when  $\langle R \rangle \gg a_0$ . It should be noted though that this unphysical finite-size effect is different from lattice type effects. Some lattice types (e.g. square lattices with four nearest neighbors) exhibit a strong lattice anisotropy, which in turn may cause the grain growth simulated using these types of lattices to stop artificially. A comprehensive review of the effects of

various neighborhood types on the kinetics of grain growth can be found elsewhere (Raabe, 2002). Other researchers attributed the low grain growth exponent in the small grain size regimes to the smaller probability of successful reorientation attempts (Holm, 1992; Radhakrishnan & Zacharia, 1995; Song et al., 1998). Therefore, the grain growth exponent varies with the grain size and asymptotically approaches a value close to 0.5 at a grain size of 15.8 when using modified MC algorithms (Radhakrishnan & Zacharia, 1995; Mehnert & Klimanek, 1997). In addition, it was argued that in the presence of second-phase particles the simulation results are significantly affected by Zener pinning (Mehnert & Klimanek, 1996; Kad & Hazzledine, 1997; Soucail et al., 1999), and furthermore, three-dimensional (rather than cross-sectional) grain size distributions should be used in some cases (Xiaoyan et al., 2000).

In view of the phenomena discussed above, large lattice systems have to be used in order to preserve a sufficient number of grains for statistical analysis since the data in small grain size regimes have to be discarded (Grest et al., 1988; Holm et al., 2001; Miodownik, 2002). However, large lattice sizes are undesirable because they hamper the application of the MC method in engineering practice. A modified MC algorithm developed later was shown to generate  $n \approx 0.5$  for both small and large grain sizes (Yu & Esche, 2003a).

In this chapter, first some theories on the self-similarity and kinetics of normal grain growth of isotropic single-phase materials are briefly reviewed. Then, both the two-dimensional CMC algorithm and various modifications and improvements introduced later are described, followed by a discussion of the extension of the MC method to three-dimensional problems. Finally, multi-scale approaches for microstructure prediction are briefly introduced.

## 2. Theories on self-similarity and kinetics of normal grain growth

### 2.1 Burke and Turnbull's analysis

For a spherical boundary with curvature  $\kappa$ , the velocity  $v$  of the boundary movement toward its center of curvature is (Burke & Turnbull, 1952)

$$v = M_0 \gamma \kappa \quad (3)$$

where  $M_0$  is the mobility and  $\gamma$  is the grain boundary energy. Assuming that for a grain of radius  $R$  the radius of curvature  $R_\kappa$  of each spherical boundary face is proportional to  $R$ , then the time rate of change of  $R$  is equal to the velocity  $v$  of the boundary movement. Therefore,

$$R_\kappa = \frac{2}{\kappa} \sim R \quad (4)$$

$$\frac{dR}{dt} = \frac{2M_0\gamma C}{R} \quad (5)$$

where  $t$  is the time and  $C$  is a dimensionless constant. Moreover, let us assume that the average grain radius  $\langle R \rangle$  evolves with the same functional relationship as any arbitrary  $R$ , which is equivalent to assuming structural self-similarity. Then, for an isotropic system (for which  $M_0$  and  $\gamma$  are constants), substituting  $\langle R \rangle$  for  $R$  into Eq. (5) and integrating this equation yields:

$$\langle R \rangle^2 - \langle R \rangle_0^2 = 2M_0\gamma Ct \quad (6)$$

However, it should be noted that in reality the entire grain boundary rarely has a uniform curvature.

## 2.2 Von Neumann's law

Von Neumann's law (Von Neumann, 1952; Mullins, 1956) is based on topological considerations. In the two-dimensional case, the rate of change of the area of a grain G is

$$\frac{dA}{dt} = -\int_G \nu ds \quad (7)$$

where A is the grain area, s is the length of the grain boundary and the integration is carried out over all grain boundary segments between grain corners where three or more grains meet. Evaluating the integral of the grain boundary curvature  $\kappa$  around a two-dimensional grain G results in

$$\int_G \kappa ds = 2\pi - \sum_{i=1}^{n_e} \theta_i \quad (8)$$

where  $\kappa$  is positive toward the center of the grain and the  $\theta_i$  are the complements of the interior angles at each of the  $n_e$  grain corners. Let us now assume that the only stable grain corner in an isotropic grain system is the three-grain junction (referred to as tri-junction) and furthermore that all grain corner angles at tri-junctions are equal to  $2\pi/3$  in order to balance the interfacial energy. Thus,  $\theta_i = \pi/3$  and for isotropic systems, the following equation can be deduced:

$$\frac{dA}{dt} = -M_0\gamma \int_G \kappa ds = -\frac{1}{3}\pi M_0\gamma (6 - n_e) \quad (9)$$

If grain growth is assumed to exhibit structural self-similarity, then by integrating the above equation, the average grain area  $\langle A \rangle$  can be shown to be proportional to the time t, and thus, the parabolic grain growth law is obtained.

## 2.3 Mullins' self-similarity theory

Consider a system with total volume V that contains a large number of particles N. Then, the mean particle volume  $\bar{V}$  may be written in the form (Mullins, 1986):

$$\bar{V} = \frac{V}{N} \quad (10)$$

Differentiating the above equation with respect to time t results in:

$$\frac{d\bar{V}}{dt} = -\frac{\bar{V}}{N} \frac{dN}{dt} \quad (11)$$

Using the continuity equation, it can be shown that (Mullins, 1986)

$$\frac{dN}{dt} = \lim_{V_0 \rightarrow 0} \int_{-\infty}^0 \dot{V} \rho(V_0, \dot{V}, t) d\dot{V} \quad (12)$$

where  $V$  is the volume of individual particles with  $\dot{V} = dV/dt$ ,  $V_0$  is the volume at  $t = 0$  and  $\rho$  is the probability density function. Therefore,

$$\begin{aligned} \frac{d\bar{V}}{dt} &= -\bar{V} \lim_{V_0 \rightarrow 0} \int_{-\infty}^0 \dot{V} f(V_0, \dot{V}, t) d\dot{V} \\ f(V_0, \dot{V}, t) &= \frac{1}{N} \rho(V_0, \dot{V}, t) \end{aligned} \tag{13}$$

where  $f$  is referred to as normalized density function. If the statistical self-similarity (SSS) holds (i.e. if  $f$  is time independent), then the following equation can be derived

$$[\bar{V}(t)]^{1-\alpha} - [\bar{V}(0)]^{1-\alpha} = (1-\alpha)Ct \tag{14}$$

where  $C$  is a constant and  $\alpha$  is a constant that depends on the system under consideration. For isothermal grain growth,  $\alpha$  is found to be  $1/3$ , and since  $\langle R \rangle$  is proportional to  $\bar{V}^{1/3}$ , the parabolic grain growth law is therefore obtained again. It should be noted that, as long as the SSS holds, this result applies to both isotropic and anisotropic grain growth.

### 3. Conventional Monte Carlo method for grain growth simulations

In the conventional MC method for simulating two-dimensional (2D) isotropic normal grain growth of single-phase materials (Anderson et al., 1984; Holm, 1992), a continuum microstructure is mapped onto a 2D lattice, whereby the most commonly used lattice type is the triangular lattice with six nearest neighbors (Rollett, 1997).

First, an integer number  $S_i$  between 1 and  $Q$  is assigned to each lattice site (where  $Q$  represents the total number of grain orientations in the system). This process is known as the initialization of the lattice. Then, the MC algorithm iteratively transforms the lattice in a procedure comprising the following four-steps: (i) random selection of a lattice site, (ii) assignment of a new orientation number to this site, which is selected randomly from all the other  $Q-1$  orientation numbers in the system, and (iii) calculation of the net change of the system energy  $\Delta E$  due to the reorientation at the selected lattice site. The total energy  $E$  is usually defined as

$$E = J_{gb} \sum_{\langle ij \rangle} (1 - \delta_{S_i, S_j}) \tag{15}$$

where  $J_{gb}$  is the grain boundary energy scale,  $\langle ij \rangle$  is the nearest neighbor site pair and  $\delta_{S_i, S_j}$  is the Kronecker delta function. Note that in certain cases it is necessary to include the second nearest neighbors into the energy calculations (Anderson et al., 1989, Holm et al., 1991). Finally, in step (iv) of the MC algorithm, the reorientation attempt is accepted with a probability  $p$

$$p(\Delta E) = \begin{cases} 1 & \text{if } \Delta E \leq 0 \\ \exp\left(-\frac{\Delta E}{kT}\right) & \text{if } \Delta E > 0 \end{cases} \tag{16}$$

where  $k$  is the Boltzmann constant and  $T$  is the simulation temperature (not the physical temperature), which is introduced to avoid stagnation of the lattice evolution in some cases (Holm et al., 1991). If  $\Delta E$  is non-positive (i.e.  $\Delta E \leq 0$ ), the attempted reorientation is accepted. If  $\Delta E > 0$ , a random number between 0 and 1 is generated and the attempted reorientation is accepted if this random number is smaller than the probability  $p$ , and otherwise, the old orientation of the site is recovered. In practice, a zero temperature probability is often adopted (Holm et al., 1991):

$$p(\Delta E) = \begin{cases} 1 & \text{if } \Delta E \leq 0 \\ 0 & \text{if } \Delta E > 0 \end{cases} \quad (17)$$

In the MC method, the temporal evolution of the simulated physical process is modeled in terms of a simulation time scale referred to as Monte Carlo Steps (MCS). Each MCS comprises  $N$  reorientation attempts, where  $N$  is the total number of lattice sites in the system. Furthermore, the MCS is assumed to be linearly related with the physical time through a jump frequency that depends on the physical temperature (Safran et al., 1983, Raabe, 2000).

In the context of the MC method, a grain is defined as a set of adjacent lattice sites that are associated with identical orientation numbers. The grain boundaries are formed by the interface lines between site pairs of unlike orientation. Then, the area  $A$  of a grain is defined as the number of lattice sites within one grain, and the radius  $R$  of a grain is commonly defined as the square root of the corresponding grain area (Holm, 1992). Lastly, the mean grain radius  $\langle R \rangle$  is defined as the square root of the total number of lattice sites divided by the total number of grains.

It should be pointed out that the underlying algorithm of the conventional MC method is stochastic in nature and thus does not reflect the physical principles inherent in the grain growth process. Its major physical and numerical shortcomings are its inability to predict the theoretically expected grain growth exponent  $n$  in the small grain size regime, the possible occurrence of unrealistic nucleation events and its general numerical inefficiency.

The normal grain growth is curvature driven and obeys a power law of the form of Eq. 1 (Hillert, 1965, Louat, 1974). For large grain radii  $\langle R \rangle$ , this relationship simplifies to the form of Eq. 2. There is ample theoretical (Burke & Turnbull, 1952; Mullins & Vinals, 1989), computational (Weiare & Kermode, 1984; Grest et al., 1988) and experimental (Weiare & Kermode, 1983; Glazier et al., 1987) evidence that the grain growth exponent equals  $n = 0.5$ . Nevertheless, a lower grain growth exponent is often reported for the earlier stages of simulations with the conventional MC method (e.g.  $n = 0.41$ , Anderson et al., 1984), which contradicts the value of  $n = 0.5$  that is theoretically expected under the assumption of self-similarity. The lower than expected grain growth exponent is usually said to be the result of the continuously decreasing boundary mobility in the early transition stage of simulated grain growth, whereas in the later simulation stages, the grain growth kinetics stabilizes (Holm, 1992; Radhakrishnan & Zacharia, 1995; Song et al., 1998). Some other researchers opined that numerical finite size effects are likely to dominate when the grain size is small (Grest et al., 1988; Anderson et al., 1989). Others suggested that the lower than expected grain growth exponent observed in the small grain size regime of the simulation is due to the mean grain size  $\langle R \rangle$  and the lattice spacing  $a_0$  representing two competing length scales (Kumar et al., 1987; Beenakker, 1988). Only if  $\langle R \rangle \gg a_0$  is the effect of the finite lattice spacing  $a_0$  negligible and the grain growth expected to be truly self-similar.

The second deficiency of the conventional MC algorithm is the occurrence of grain nucleation by reorientation in some instances, for example in simulations with non-zero temperature and in simulations of anisotropic grain growth. Of course, this nucleation phenomenon does not correlate with the actual physical grain growth process. In non-zero temperature simulations, a grain might nucleate at any lattice site due to the artificially introduced "thermal fluctuation", in spite of this nucleation causing an increase in the system energy. Also, if the grain boundary energy is anisotropic, a new grain with smaller grain boundary energy may be nucleated by reorientation alone – a phenomenon that in reality would be prevented based on thermodynamic considerations.

The third shortcoming of the conventional MC method is that it is very time consuming (Bortz et al., 1975; Song et al., 1998). This inefficiency is attributable mainly to the fact that the vast majority of attempted lattice point reorientations cause net energy increases, and therefore, the probability for energy-reducing reorientations leading to actual net grain growth is extremely small. In addition, for large two-dimensional (2D) or three-dimensional (3D) lattice systems and for large numbers of possible orientations  $Q$ , this situation even worsens progressively.

Scaling constitutes another important problem associated with the MC method since it does not involve any physical time or length parameters. The only time constant involved is the MCS, and the spacing of the simulation lattice represents the only length constant. Thus, in order to relate the MC simulation results to actual physical units, both time and length scaling is needed. While several scaling approaches for the MC method have been suggested, this issue is still far from being resolved (Raabe, 2000; Raabe, 2002; Janssens et al., 2007; Nosonovsky et al., 2009).

#### 4. Modified Monte Carlo algorithms for grain growth

From a purely computational point of view, the  $n$ -fold method, which was originally proposed for the Ising model (Bortz et al., 1975), was later extended to zero-temperature grain growth simulations (Sahni et al., 1983). In the  $n$ -fold method, the lattice sites were separated into two types. The first type consists of those lattice sites for which any orientation switches are rejected because they increase the total energy in the system. The second type comprises those sites for which orientation switches to at least one of the other  $Q-1$  orientations cause a reduction in the total energy. Later, the concept of site activity – a parameter that is calculated based on the orientations of the neighboring sites – was introduced (Holm, 1992). In this approach, a site of the second type is randomly selected with a frequency proportional to its activity, and a new orientation is selected randomly among the energetically favorable orientations. One additional random number is used to estimate the time elapsed (measured in MCS) between two successive reorientation attempts. This method ensures that every reorientation attempt is successful, and therefore it significantly increases the computational efficiency. However, there are four major shortcomings associated with the  $n$ -fold method. First, the algorithm is more complicated than other modifications of the MC method. Second, an additional random number is required that has no clear physical meaning. Third, this technique is only valuable and applied in the later stages of grain growth when most of the lattice sites are bulk sites and cannot be reoriented successfully. Thus, the grain growth kinetics obtained in the early stages of the simulation is not affected. Fourth, artificial grain nucleation and grain coalescence may still occur in some cases of non-zero temperature simulation or anisotropic

grain boundary energies, and therefore this method may not be universally applicable without further modifications.

The early MC algorithms for modeling grain growth did not incorporate the physics of the process. Later, a modified MC method that considers the grain growth physics was suggested (Radhakrishnan & Zacharia, 1995), although the resulting algorithm is similar to the n-fold method discussed above. It was argued that, since in reality grain growth occurs through atom migration from one grain to another across the corresponding grain boundary, an attempted reorientation should not be selected from all possible values of orientations but rather it should be restricted to one of the nearest neighbors' orientations. Therefore, it was proposed to generate a random number in the range between one and the total number of nearest neighbors of the lattice site under consideration. Based on this random number, the corresponding neighbor's orientation is assigned to the attempted lattice site for evaluation. This modification prevents artificial grain nucleation during grain growth and thus renders the MC model more realistic. Another recommendation was to select the new orientation amongst those of the unlike neighbors without weighting the selection by the number of neighbors that possess this new orientation as in the method above (Holm & Battaile, 2001). This approach is equivalent to the n-fold method.

Furthermore, it was suggested that there is no physical meaning to randomly selecting lattice sites for reorientation (Song et al., 1998; Song & Liu, 1998), and it was therefore proposed to evaluate them one by one in each MCS. By this approach, it is not only ensured that computing time is saved but also that all lattice sites are attempted for reorientation exactly once in each MCS.

The above modifications contributed to the resolution of the shortcomings associated with the conventional MC algorithm for grain growth simulations. However, they did not fully take the grain growth physics into account. Therefore, they generated grain growth exponents close to  $n = 0.5$  only for very large average grain sizes  $\langle R \rangle$ , while the lower grain growth exponents typically obtained in the early simulation stages remained basically unchanged. This imperfect modeling may be attributable to the procedural algorithm design methodology employed. More recently, an algorithm that better reflects the physical behavior observed in nature was introduced (Yu & Esche, 2003a). The accuracy and efficiency of an algorithm may be affected by its design methodology, whereby procedural and object-oriented approaches represent the two fundamental methodologies. The object-oriented design is preferable for modeling the static and dynamic behaviors of physical objects without using analytical equations of motion (Yu & Esche, 2004). Note that the conventional MC algorithm for simulating the grain growth process was designed in a procedural fashion. While it was able to reproduce some essential process features with relative success, it still exhibited significant potential for improvements in accuracy and efficiency. The n-fold method discussed above significantly reduced the computational time compared with corresponding MC algorithms. However, it did neither conceptually change the conventional MC model nor improve the algorithm's accuracy. The algorithms proposed thereafter started to take into consideration the grain growth physics, and therefore, the efficiency, accuracy and capabilities of the MC simulations were greatly improved. However, these algorithms were still designed in a procedural manner and missed some essential physical considerations. Facilitated by the object-oriented algorithm design methodology, three modifications to the conventional MC algorithm were proposed (Yu & Esche, 2003a).

**4.1 Modification I**

From the point of view of object-oriented programming, the lattice is composed of grains, and the lattice sites forming a grain are either at the boundary or in the interior of that grain. Grain growth occurs through grain boundary migration, or in other words, through the jump of atoms from one grain to another through the grain boundary. In the modified MC algorithm (Yu & Esche, 2003a), when a lattice site is selected, its location is checked first. Interior sites do not migrate. Therefore, no reorientation is tried for such sites. If the selected site is located at a grain boundary, then the algorithm attempts a possible reorientation.

In order to save computing time, it was earlier proposed not to check for reorientation those lattice sites for which more than half of the neighboring lattice sites were of the same orientation (Sahni et al., 1983). This method does not exhibit any problems for isotropic systems. However, in systems with significant anisotropy of the grain boundary energies, this approach might under certain circumstances lead to the exclusion of viable, energy-reducing reorientation attempts for those lattice sites. This method is not in accordance with the physics of grain boundary movements, and therefore it cannot be applied universally.

**4.2 Modification II**

Consider the grain structure in Figure 1. Suppose that the lattice site under consideration (circled) is located at the boundary of the grain with orientation 3 (grain 3 for short). The grains adjacent to this lattice site have orientations of either 1 or 2. Furthermore, let us assume that grain 3 is currently shrinking. In that case, its boundary sites would either jump to the neighboring grains 1 or 2 or keep their current orientation, depending on whether the energy criterion for reorientation is satisfied or not. Thus, the new attempted orientation for the lattice site under consideration for reorientation should be 1 (or 2) if the corresponding site is adjacent to grain 1 (or 2) only or if the site is a triple point where three grains meet. In either case, the selection of the new orientation should be limited to those of the neighboring grains.

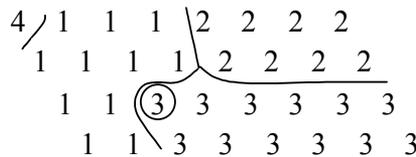


Fig. 1. Schematic of a grain structure

Modification II to the conventional MC algorithm (Yu & Esche, 2003a) involves the generation of a random number in the range between 1 and q, where q is the total number of nearest neighboring sites with orientations that differ from that of the attempted boundary site. Then, this random number is utilized to pick one of these neighbors and to assign its orientation to the attempted site. The improvement of this modification over the one proposed earlier (Radhakrishnan & Zacharia, 1995) is that using Modification II, the circled site of grain 3 in Figure 1, for instance, would definitely jump to grain 1 and thus reduce the system energy. With the previously proposed modification, though, it would have jumped only with a probability of 4 out of 6. Comparing Modification II with the prior suggestion (Holm & Battaile, 2001), the selection of the new orientation is weighted by the total number of neighboring lattice sites with the same orientation. This is probably the case in a physical sense since grain growth represents a free energy reduction process that involves random

movements to some degree, and very possibly, in the physical reality the free energy reduction occurs following the shortest possible path. Therefore, this modification provides the boundary sites with an optimal mobility and represents a more realistic grain growth model than the conventional MC algorithm as will be shown by the sample simulations below.

### 4.3 Modification III

When using the conventional MC algorithm, a subset of all lattice sites is considered multiple times for possible reorientation while some of the remaining sites are not attempted at all. Consider for instance a  $200 \times 200$  lattice, introducing an index numbering the lattice sites from 1 to 40,000. There are two possible methods for generating the random number needed to select a lattice site whereby the random integer numbers can be generated using the function `rand()` in the standard C library. Method 1 is  $40,000 \times \text{rand()} / \text{RAND\_MAX}$ , where `RAND_MAX` = 32767 is the maximum integer value that can be generated by `rand()`. Method 2 consists of producing two integer random numbers in order to obtain the coordinates of the lattice site to be selected for possible reorientation and then calculating the index of that site. In both methods, only a small fraction of all possible site indices occur exactly once in the 40,000 reorientation attempts while a significant portion do not occur at all (Yu & Esche, 2003a). In each MCS, additional random numbers are also generated to select new orientations, but this is not likely to affect the uneven selection of lattice sites. As a consequence, the objective of attempting reorientation for every lattice site in each MCS cannot be met in a completely random fashion. Therefore, the microstructure does not evolve evenly in the entire lattice domain, and thus, the simulated grain growth kinetics is likely to be affected.

For that reason, another modification that eliminates the possibility of multiple selections of one lattice site within one MCS was introduced into the conventional MC algorithm (Yu & Esche, 2003a). In this approach, the first lattice site is selected randomly from the total of  $N$  lattice sites in the system and evaluated for possible reorientation. Then, the second site is selected randomly only from the remaining  $N-1$  lattice sites and then the next ones amongst the remaining  $N-2$ ,  $N-3$  and so on. In this fashion, all sites are selected exactly once per MCS. Note that this modification still involves some element of randomness, which can be justified by two physical reasons. First, a grain boundary moves towards its center of curvature in a somewhat random fashion at smaller length scales. In fact, it was argued earlier that the motion of the grain boundary could be considered as sections of the boundary undergoing a random walk (Louat, 1974). Second, it would be desirable to update all boundary locations at the same time, but a method to achieve that objective without causing evolution stagnation has not been devised yet. On the other hand, when implementing Modification III, the movement of all grain boundaries during each MCS can be regarded as an approximation to the simultaneous movement of all boundaries at one physical time instant.

The above-mentioned earlier modification where all lattice sites are evaluated sequentially in each MCS (Song et al., 1998) might not correctly model the natural behavior of grain growth. In fact, when combining this approach with certain selection methods for the candidate orientations, it does not work at all. This can be demonstrated for instance by simulating the shrinkage of a hexagonal grain that is imbedded in an infinite matrix using a triangular lattice as shown in Figure 2. When combining this case with Modification II discussed above using zero-temperature probability and selecting reorientation sites one by

one from left to right and top to bottom, the imbedded grain disappears immediately in a single MCS. Of course, this outcome does not reflect the correct grain shrinkage kinetics. Other conceivable artificial site selection algorithms may exhibit similar problems, and therefore, they would be undesirable if they lack underlying physical meaning.

```

0 0 0 0 0 0 0 0
  0 0 0 0 0 0 0 0
    0 0 0 1 1 1 0 0
      0 0 1 1 1 1 0 0
        0 0 1 1 1 0 0 0
          0 0 0 0 0 0 0 0
            0 0 0 0 0 0 0 0

```

Fig. 2. A hexagonal grain (1) imbedded in a matrix (0)

#### 4.4 Simulation results

As the simulation results and the discussion below demonstrate, the differences between the three modifications to the conventional MC algorithm and other modifications suggested previously are especially critical when considering small grain size regimes, where a large fraction of the lattice sites are located at grain boundaries. A detailed analysis of the performance of the three modifications to the conventional MC algorithm has been conducted by simulating grain shrinkage and grain growth (Yu & Esche, 2003a), in which the resulting grain growth exponents  $n$  obtained using least square regression analysis and the total number of MCS to achieve a certain mean grain radius  $\langle R \rangle$  were compared with several preexisting algorithms. Furthermore, the variability of the simulation results with respect to different seeds used by the random number generator was analyzed.

##### 4.4.1 Simulation of grain shrinkage

Simulating the kinetics of the shrinkage of an isolated circular grain that is imbedded in an infinite matrix is generally regarded as an effective means for testing the correctness and efficiency of a grain growth simulation procedure. In order to test the three modifications discussed above in the large grain size regime, the kinetics of the shrinking of a circular grain with an initial grain radius of  $R_0 = 50$  was examined employing a  $200 \times 200$  triangular lattice with six nearest neighbors. The following simulation procedures were compared:

1. Procedure 1: conventional MC algorithm
2. Procedure 2: MC algorithm with modification by Radhakrishnan and Zacharia, 1995
3. Procedure 3: MC algorithm with Modification II discussed above
4. Procedure 4: MC algorithm with Modifications II and III discussed above

The Ising Model (i.e.  $Q = 2$ ) was applied for the circular grain shrinkage simulation by Procedure 1. Since Modification I saves about 20% of the computing time without affecting the lattice evolution, this modification was adopted for all four procedures.

Subsequently, the normal grain growth up to a mean grain radius of  $\langle R \rangle = 25$  was simulated employing the four procedures listed above in conjunction with a  $400 \times 400$  lattice. Lattices of this large size were used in order to prevent the simulated grain growth kinetics from getting distorted by lattice boundary effects. For Procedures 2, 3 and 4, the total number of possible orientations  $Q$  can theoretically be equal to the total number of

lattice sites  $N$  without seriously affecting the computing time. In this study,  $Q = 10,000$  was employed for these three procedures, which is deemed large enough to render any possible artificially introduced grain coalescence insignificant. Note that for Procedure 1, the number of possible orientations has to be reduced significantly, and  $Q = 63$  was applied in this study. The affect of the choice of  $Q$  on the simulation results when using Procedure 1 is well documented in the literature. Since the major concern is in the early stages of the simulations, the  $n$ -fold method was not applied with Procedure 1.

Figure 3 shows the temporal evolution of the grain shape during shrinkage of a circular grain with an initial size of  $R_0 = 50$  imbedded in an infinite matrix as simulated using Procedure 4. As is to be expected with lattice anisotropy, the grain maintains an approximately circular shape only within statistical fluctuations.

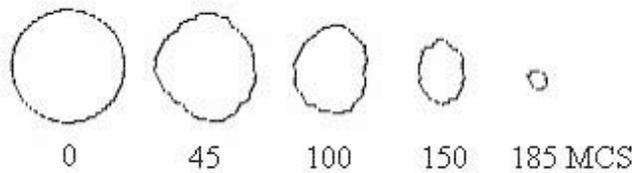


Fig. 3. Temporal evolution of circular grain ( $R_0 = 50$ ) in infinite matrix for Procedure 4

In Figure 4, the evolutions of grain area  $A$  and grain radius  $R$  are depicted as functions of time for the simulation as described above. As is expected from theory, grain area  $A$  was found to decrease linearly with time. A regression analysis was used to fit the grain radius  $R$  vs. time  $t$ , producing the following result:

$$R = 3.5109(189.4420 - t)^{0.5097} \quad (18)$$

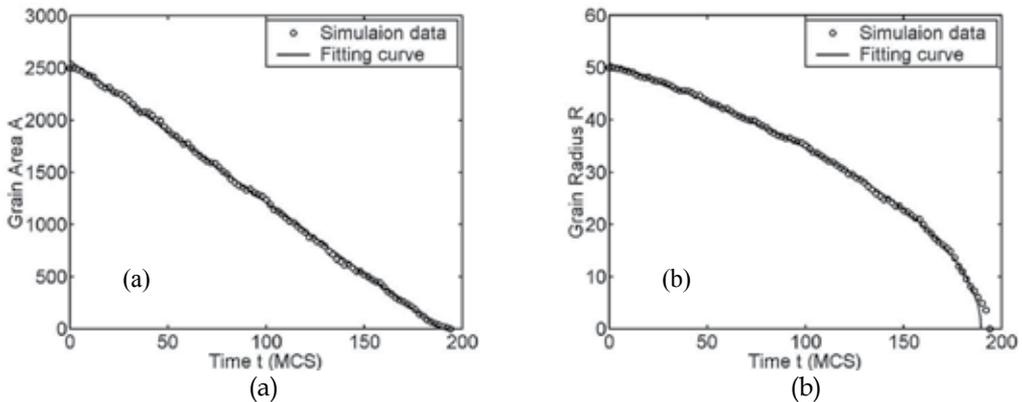


Fig. 4. Temporal evolutions of (a) grain area and (b) grain radius of circular grain ( $R_0 = 50$ ) for Procedure 4

A comparison of the grain area evolutions simulated using Procedures 1 (with  $Q = 2$ ), 2, 3 and 4 is presented in Figure 5. In all four cases, an approximately linear relationship between grain area  $A$  and time  $t$  was obtained. Also, regression analyses on data for grain radius  $R$  vs. time  $t$  yielded a grain shrinkage exponent of around 0.5 as well, which confirms

corresponding findings reported elsewhere (Anderson et al., 1984; Radhakrishnan & Zacharia, 1995). This result implies that the essential feature of curvature-driven grain growth is adequately captured by the MC.

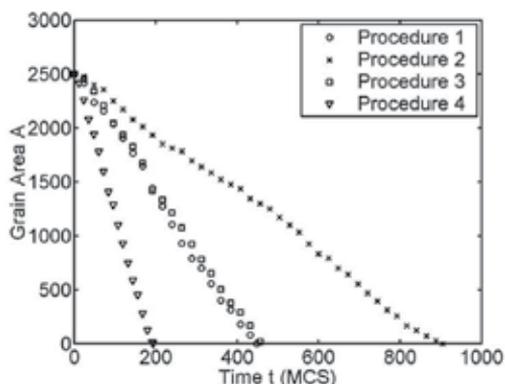


Fig. 5. Temporal evolution of circular grain area  $A$  for Procedures 1 through 4

It should be pointed out that Procedure 4 causes the grain to disappear in the shortest time. Furthermore, a comparison of the simulation results obtained with Procedures 3 and 4 reveals that the modified approach for selecting candidate lattice sites for reorientation (Modification III) reduced the number of MCS by roughly 50%. Also, the area evolutions for Procedure 1 (with  $Q = 2$ ) and Procedure 3 are identical within the expected margins of error because these two procedures are essentially equivalent. Finally, a 50% reduction in MCS compared with Procedure 2 is achieved by employing Modification II for selecting candidate new orientations.

#### 4.4.2 Kinetics of normal grain growth

Next, the computational efficiency of simulating isotropic normal grain growth using Procedure 4 was studied (Yu & Esche, 2003a). Figure 6 depicts the microstructures simulated using the modifications discussed above at two instances. It was confirmed by careful examination of the microstructures obtained in the simulations that small grains tend to shrink and ultimately disappear while large grains grow, which is in accordance with simulation results reported previously (Srolovitz et al., 1984a). The morphology of the obtained grain structure is compact and similar to the real material microstructure. Some unrealistic straight grain boundaries are observable in Figure 6b though, which are attributable to the anisotropy of the lattice used (Holm et al., 1991). Such straight boundaries may be eliminated by either artificially introducing noise into the system in the form of a non-zero temperature probability or by including the second nearest neighbors into the energy calculations. However, the simulation temperature is an artificial parameter, and thus, its inclusion is not desirable from the perspective of object-oriented algorithm design. Also, while including the second nearest neighbors into the energy calculations appears physically justifiable, the relative weights assigned to the contributions of first and second nearest neighbors may again be considered as artificial parameters that are objectionable from the point of view of object-oriented algorithm design. Further developments on the construction of lattices with less pronounced anisotropy would therefore be desirable (Yang et al., 1995).

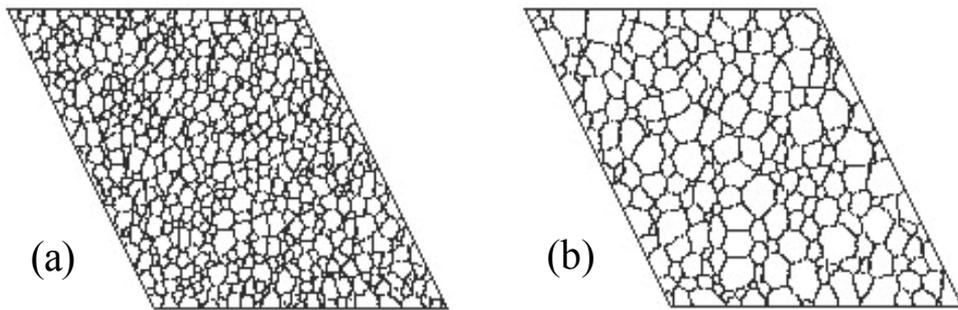


Fig. 6. Isotropic normal grain growth by Procedure 4 after (a) 100 MCS, (b) 285 MCS

Typical results for temporal evolutions of the mean grain radius obtained using the four procedures are shown in Figure 7. The grain growth exponents were calculated from Eq. 2 as the slopes of the log-log plots for grain radius vs. time, which is common practice (Anderson et al., 1984; Holm, 1992; Radhakrishnan & Zacharia, 1995; Song et al., 1998). The data points for  $\langle R \rangle < 7$  were excluded from the calculations of the grain growth exponents because they deviate from the linear relationship between  $\log(\langle R \rangle)$  and  $\log(t)$ . This deviation may be attributed to the effect of the disorderly lattice structure early in the simulation and the neglecting of the initial grain radius  $R_0$  in Eq. 2. The corresponding grain growth exponents for the four procedures (averaged over 10 runs) were reported elsewhere (Yu & Esche, 2003a).

The grain growth exponents were found to be  $n = 0.5$  in the large grain size regimes for all four procedures, which is in accordance with earlier reports (Grest et al., 1988; Radhakrishnan & Zacharia, 1995). However, only the simulations using Procedure 4 generated  $n = 0.5$  in the earlier stage ( $\langle R \rangle < 15$ ), and the simulations using Procedure 3 resulted in values of  $n$  between 0.44 and 0.49. Therefore, both Modifications II and III affect the grain growth kinetics.

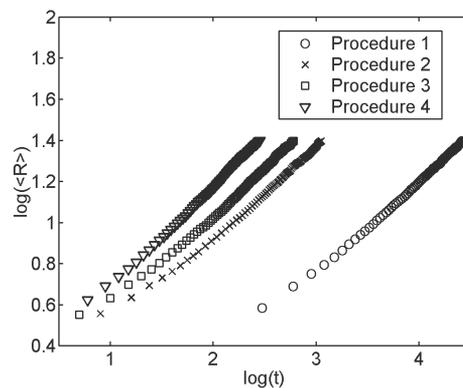


Fig. 7. Grain growth kinetics for all four procedures

#### 4.4.3 Grain growth exponent

Since a grain growth exponent of  $n = 0.5$  was obtained in the large grain size regimes using the conventional MC algorithm with the  $n$ -fold method (Grest et al., 1988), it was then hypothesized that unrealistic (unphysical) finite size effects are likely to dominate in the

early simulation stages, which would in turn lead to lower values of  $n$  in the small grain size regimes. However, the modified MC algorithm discussed above generates  $n = 0.5$  even in the early simulation stages. The obtained growth exponent was further validated by fitting the data for average grain radius  $\langle R \rangle$  vs. time  $t$  to Eq. 1. A typical result is given in Figure 8.

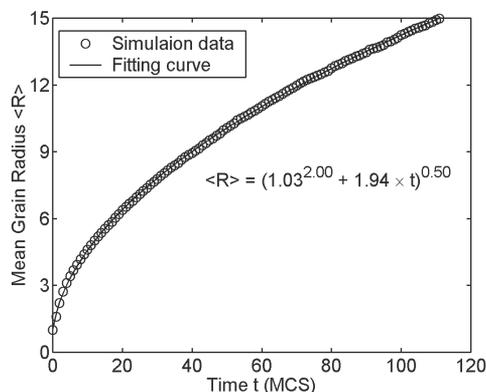


Fig. 8. Regression analysis for data using Procedure 4 with objective function  $\langle R \rangle^m - \langle R \rangle_0^m = Mt$  with parameters  $m$ ,  $\langle R \rangle_0$  and  $M$  to be determined

By analyzing the distributions of the normalized grain sizes and the numbers of edges per grain, the grain evolution was found to be a self-similar process (see Figure 9). From Figure 9b, it is evident that the grain shapes remain essentially unchanged, with the exception of the very beginning of the simulation ( $t < 1$  MCS). This means that, when modeled using Procedure 4, the grain boundary movement immediately results in a steady state, despite the fact that the initial lattice structure is random. Additional support for this observation is provided by the value of  $\langle R \rangle_0 = 1.03$  obtained by regression analysis (see Figure 8). Therefore, it appears that Procedure 4 leads to the expected parabolic growth law, independently of any size effects. This conclusion is further supported by a value of  $n = 0.5$  obtained in the Ising model using the conventional MC algorithm (Anderson et al., 1984; Grest et al., 1988). Note that in this case the modeling of the boundary movement is equivalent to Modification II discussed above. Furthermore, there is no theoretical basis for a dependence of the grain growth exponent on the grain size. In fact, in Mullins' theoretical deduction of the parabolic growth law under the assumption of statistical self-similarity (Mullins, 1986), the grain size was not taken into consideration as long as there are enough domains in the system in order to avoid surface effects and if the material volume is differentiable, i.e., the material is continuous at the length scale of the grain size considered. Even though the lattice in the MC method represents a set of discrete points, it is the continuous grain boundaries whose movements are modeled in the MC algorithm. Thus, a continuous material can be mapped onto the lattice without loss of the continuum property. Therefore, the MC lattice is appropriate for representing the theoretical material model employed by Mullins, and the MC algorithm can be expected to reproduce the theoretical result. Note that it indeed does, as is documented through the results discussed above.

The reasoning above demonstrates that the lower grain growth exponents reported for the conventional MC algorithm cannot be explained by finite size effects. The temporal evolutions of the normalized grain size and grain shape distributions obtained using the conventional MC algorithm are self-similar (Srolovitz et al., 1984a) and resemble those shown in Figure 9. A

careful study of the growth kinetics obtained using the conventional MC algorithm suggested that its strong random nature significantly affects the growth kinetics and the results of the regression analysis used to fit the simulation data (Yu & Esche, 2003b). By excluding the data from the very early stages, where the randomness effects are likely to dominate the simulated lattice evolution, and by employing three-parameter regression analyses based on Eq. 1, growth exponents close to  $n = 0.5$  can be obtained.

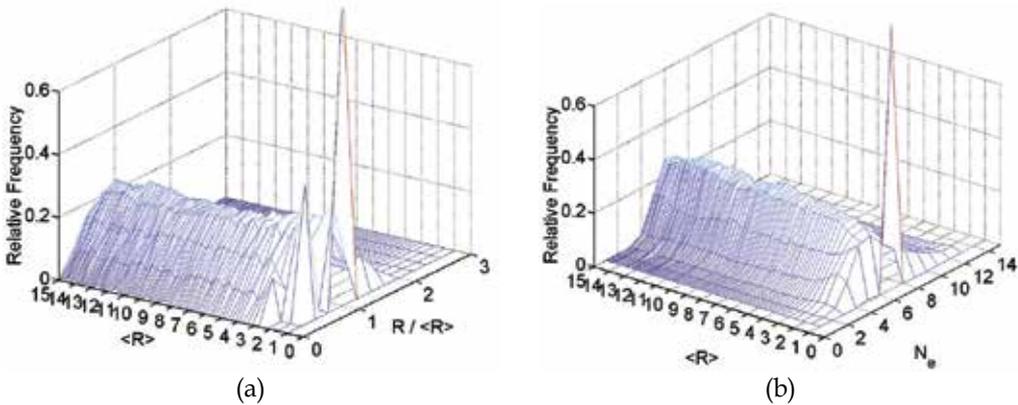


Fig. 9. Distributions for Procedure 4 of (a) grain radius  $R$  normalized by mean grain radius  $\langle R \rangle$  and (b) number of edges per grain  $N_e$

While the lower grain growth exponent measured experimentally for zone-refined materials was attributed to the initial grain morphology and grain size distribution among other reasons (Humphreys & Hatherly, 1996), the parabolic grain growth kinetics still occurs even for unstructured initial lattices as is shown in Figure 9. Therefore, the MC simulations do not support this reasoning.

#### 4.4.4 Algorithm efficiency

Both the total number of MCS and the total computing time to reach a certain grain size could be considered as parameters that characterize the computational efficiency of the MC algorithm. An investigation of the computational efficiency of the four alternative procedures discussed above was conducted using a  $400 \times 400$  triangular lattice (Yu & Esche, 2003a). While the average numbers of MCS per second were found to be similar for all four procedures, the total numbers of MCS and the total computing times for Procedure 4 were significantly lower than those for the remaining three procedures. Thus, according to both measures, Procedure 4 exhibits by far the best efficiency.

## 5. Three-dimensional Monte Carlo method for grain growth simulations

### 5.1 Background

The two-dimensional grain growth of ideal single-phase materials has been studied extensively both by theoretical deductions (Burke & Turnbull, 1952) and by computer simulations (Anderson et al., 1984; Holm & Battaile, 2001; Yu & Esche, 2003a). However, the grain growth process represents a three-dimensional (3D) phenomenon. While the generalization of von Neumann's law from 2D to 3D has been attempted (Rivier, 1983; Liu

et al., 2001), these 3D models are still not widely accepted and necessitate further development (Mullins, 1986; Atkinson, 1988). In light of the technical difficulties associated with experimental studies of 3D grain growth, computer simulation represents an effective tool for exploring this 3D microstructural event. A number of Monte Carlo (MC) Potts models for the 3D grain growth process have been developed (Anderson et al., 1989; Song & Liu, 1998; Sista & DebRoy, 2001). In these 3D simulations, the grain growth exponent  $n$  was found to either asymptotically approach the theoretical value in the long-time limit (Radhakrishnan & Zacharia, 1995) or to be significantly lower than the theoretically expected value over the entire time domain (Yang et al., 2000). Furthermore, the normalized grain size distributions obtained in 3D MC Potts simulations was time-dependent (Xiaoyan et al., 2000). In addition, the 3D lattices employed in these earlier 3D simulations were usually limited to a size of  $100 \times 100 \times 100$ . Therefore, these simulations had to be terminated at relatively small grain sizes in order to preserve adequate numbers of grains for statistically significant analysis and could not adequately support the theoretical work. A modified MC Potts algorithm that had been shown to produce the theoretically expected grain growth exponent over the entire time domain in 2D MC Potts simulations (Yu & Esche, 2003a; Yu & Esche, 2003b) was later expanded to model 3D grain growth (Yu & Esche, 2003c; Yu et al., 2005) and will be briefly described here.

## 5.2 Implementation

In the 3D MC Potts algorithm for isotropic, single-phase grain growth simulation, a continuum microstructure was mapped onto a  $200 \times 200 \times 200$  cubic lattice with 26 nearest neighbors (i.e. including third nearest neighbors). A MCS was defined as  $N$  reorientation attempts, where  $N = 8,000,000$  here. The radius  $R$  of a 3D grain was defined as the cubic root of the grain's volume (i.e. the total number of lattice points within the grain). The average grain volume was defined as the total number of lattice points (i.e., 8,000,000) divided by the total number of 3D grains in the lattice, and the average 3D grain radius  $\langle R \rangle$  was defined as the cubic root of the average grain volume. The radius of a grain in a cross section was defined as the square root of the number of lattice points within the grain in the cross section, and the average grain radius for a cross section was defined as the square root of the total number of lattice points in the cross section (i.e., 40,000) divided by the total number of grains in the cross section.

The cubic lattice was initialized by randomly assigning an integer orientation number to each lattice point. Initially, the number of different orientations was  $Q = N = 8,000,000$ . In the reorientation procedure of the modified MC Potts algorithm, all lattice sites were randomly attempted for reorientation exactly once in each MCS. In the modified algorithm, the location of the lattice sites were checked first, and only the sites located at the grain boundaries were selected for a possible reorientation. For each attempted lattice site, the new orientation was selected randomly from those of its neighbors with unlike orientations. If the system energy did not rise due to the attempted reorientation, the new orientation number was accepted. Otherwise, the old orientation number was recovered. The simulations were terminated at time  $t = 250$  MCS. At that point, a grain size of  $\langle R \rangle \approx 15$  was achieved and a large number of grains were preserved in the lattice. Each simulation was repeated ten times using different seeds for the random number generation. The microstructures were examined every 5 MCS.

### 5.3 Simulation results

In order to facilitate the presentation of the simulation results, a Cartesian coordinate system was attached to the cubic lattice. For instance,  $x = 200$ ,  $y = 200$  and  $z = 200$  characterized the front, right and top surfaces of the cubic lattice, respectively, while  $z = 100$  represented the middle cross sectional plane in  $z$ -direction. Figure 10 depicts the 3D microstructure observed at the time instant when  $\langle R \rangle \approx 15$ . Figure 11 shows the temporal evolution of the microstructure in the plane with  $z = 100$ . These figures indicate that a compact grain structure was developed. Also, various microstructural features of normal grain growth commonly found in 2D MC Potts simulations were observed in the cross sections of the 3D lattice. For instance, generally small grains shrank while large grains grew up, and 120-degree angles were found at most grain corners. These observations are in accordance with findings reported earlier (Anderson et al., 1989, Song & Liu, 1998).

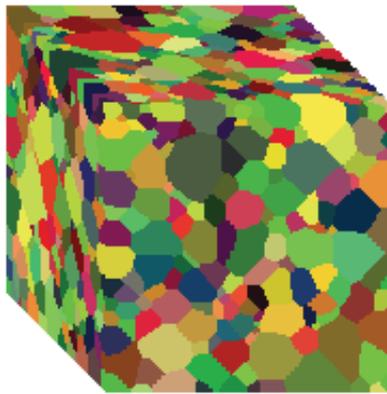


Fig. 10. 3D microstructure at 250 MCS,  $\langle R \rangle \approx 15$

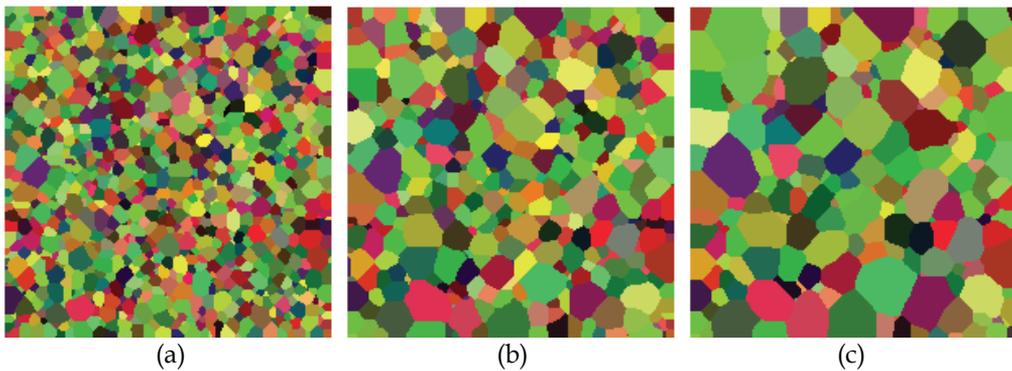


Fig. 11. Temporal evolution of microstructure in plane  $z = 100$ , (a) 50 MCS ( $\langle R \rangle_{z=100} \approx 6.4$ ), (b) 150 MCS ( $\langle R \rangle_{z=100} \approx 10.1$ ) and (c) 250 MCS ( $\langle R \rangle_{z=100} \approx 13.7$ )

The temporal evolution of the distribution of 3D grain sizes  $R$  normalized by average grain radius  $\langle R \rangle$  is illustrated in Figure 12a. It can be observed that the grain size distribution did not change while  $\langle R \rangle$  increased from 4 to 15. Also, the grain size distributions obtained in ten simulation runs were very similar, i.e., the seeds used in the random number generation influenced the grain size distribution only very insignificantly. Two alternative equations -

the Louat distribution in the form of Eq. 19 and the log-normal distribution in the form of Eq. 20 - were then fitted to the averaged grain size distribution data for the ten runs

$$F = C_1 R_n \exp(-C_2 R_n^2) \tag{19}$$

$$F = C_3 \exp[-C_4 (\ln R_n - C_5)^2] - C_6 \tag{20}$$

where  $F$  is the relative frequency,  $R_n = R / \langle R \rangle$  is the normalized grain size, and  $C_1 - C_6$  are the fitting parameters. Figure 12b summarizes the results of the regression analyses. It is apparent that the log-normal distribution provided for a better fit for the 3D simulation data than the Louat distribution. This conclusion confirms prior observations reported elsewhere (Radhakrishnan & Zacharia, 1995). Specifically, the simulations resulted in more small grains and fewer large grains than in the log-normal curve. This deviation indicates that the grains in the simulations had a higher potential for later growth - a phenomenon that had also been documented before in the literature (Srolovitz et al., 1984a).

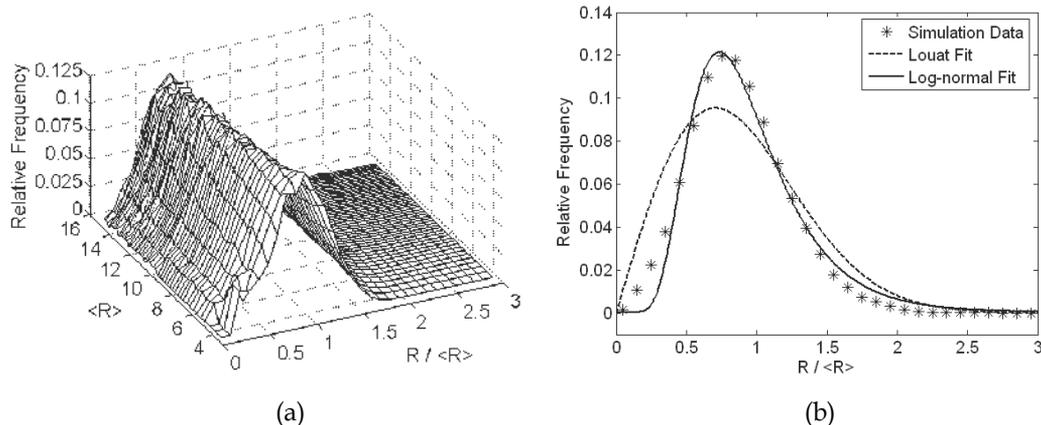


Fig. 12. (a) Time-invariant distribution of 3D grain sizes  $R$  normalized by average grain sizes  $\langle R \rangle$ ; (b) regression analyses of grain size distribution averaged over time and over 10 simulation runs

Time-invariant grain size distributions were also observed for the cross sections. A typical example is given in Figure 13a. While the grain size distributions for the individual cross sections obtained in the ten simulations exhibited relatively large variations, the averaged grain size distributions over the ten runs for the three cross sections with  $x, y, z = 100$ , respectively, differed only slightly. Regression analyses of the grain size averaged over the three cross sections showed that the simulation data can be fitted to both the Louat and the log-normal distributions (see Figure 13b). Contrary to the corresponding 3D case, the Louat distribution appears to provide a better fit, though. The fact that the Louat distribution provides a better fit for 2D simulation results had also been observed before (Anderson et al., 1989). Note also that alternative grain size distributions were discussed elsewhere (Mullins, 1991; Mullins, 1998).

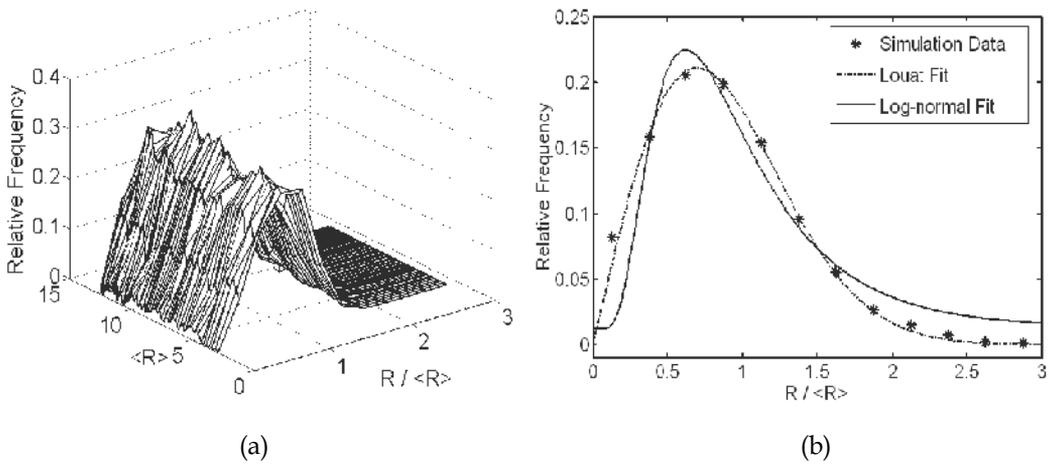


Fig. 13. Normalized grain size distributions for cross sections: (a) temporal evolution of grain size distribution in one cross section; (b) regression analyses of grain size distribution averaged over time, 10 simulation runs and 3 cross sections

Figure 14 shows the distributions of the number of facets per grain  $N_f$  and the number of edges per grain  $N_e$  for the cross sectional plane with  $z = 100$ . The  $N_f$  distribution depicted in Figure 14a appears to be statistically time-invariant, though, except for the very late simulation stage that exhibits a sharp peak. The peak values of  $N_f$  fall in the relatively narrow range between 11 and 14 as seen in Figure 15, which compares with a value of  $N_f = 13.7$  for the average number of facets per grain reported elsewhere (Krill & Chen, 2002). Furthermore, a range of 11.16 to 15.54 was reported for experimental measurements of  $N_f$  (Krill & Chen, 2002). Similarly, the distribution of edges per grain  $N_e$  is also statistically time-invariant as is seen in Figure 14b.

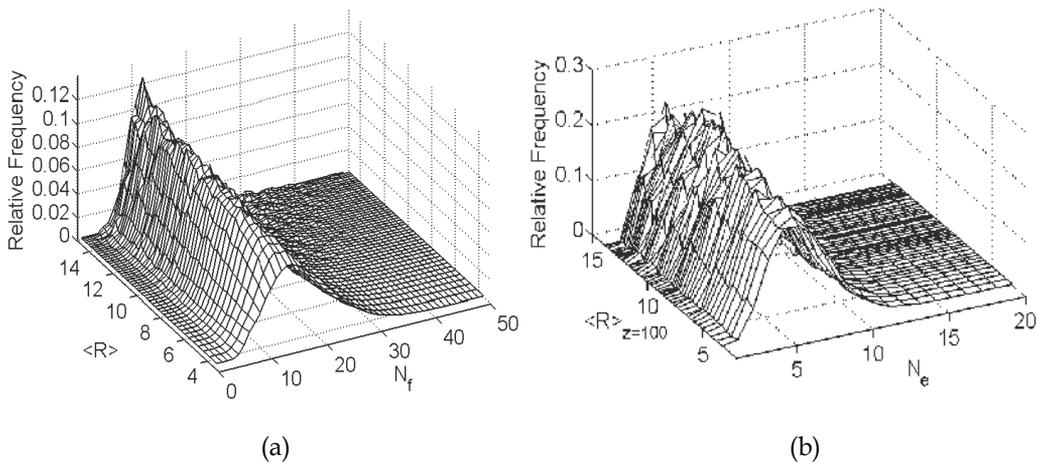


Fig. 14. Distribution of number of (a) facets per grain  $N_f$  for 3D domain and (b) edges per grain  $N_e$  for cross sectional plane with  $z = 100$

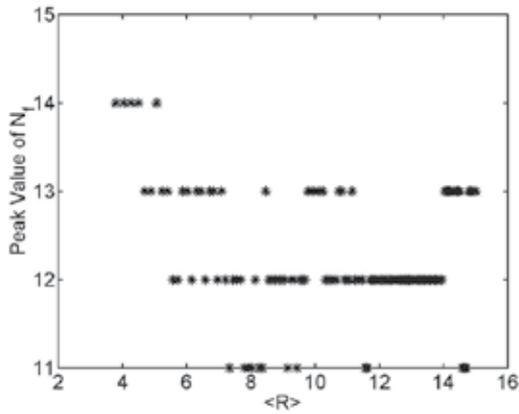


Fig. 15. Evolution of peak value of  $N_f$  (average value 12.25)

Least-square regression analyses of the grain growth kinetics were then performed based on the data obtained from the 3D domain and the three cross sections at  $x = 100$ ,  $y = 100$  and  $z = 100$ . In each case, the data for grain size vs. time were averaged over ten simulation runs prior to conducting the regression analysis. Note that the differences in grain size vs. time data between the ten simulation runs were very small. Figure 16 illustrates the data for averaged grain size vs. time and regression analysis results based on the classical form of the grain growth kinetics (i.e. Eq. 1 with fitting parameters  $\langle R \rangle_0$ ,  $m$  and  $M$  and defining the grain growth exponent as  $n = 1/m$ ). A value of  $n \approx 0.5$  was obtained both for the 3D domain as well as for the cross sections. The grain growth rates  $M$  obtained from the three cross sections were similar to each other with an average value of  $M \approx 0.74$ . Also, they were smaller than the rate of  $M \approx 0.96$  found for the 3D domain (see Figure 16b). In Figure 17, a log-log plot of average grain radius  $\langle R \rangle$  vs. time  $t$  is provided. Note that the same grain growth kinetics was also observed in simulations with grain sizes larger than 15.

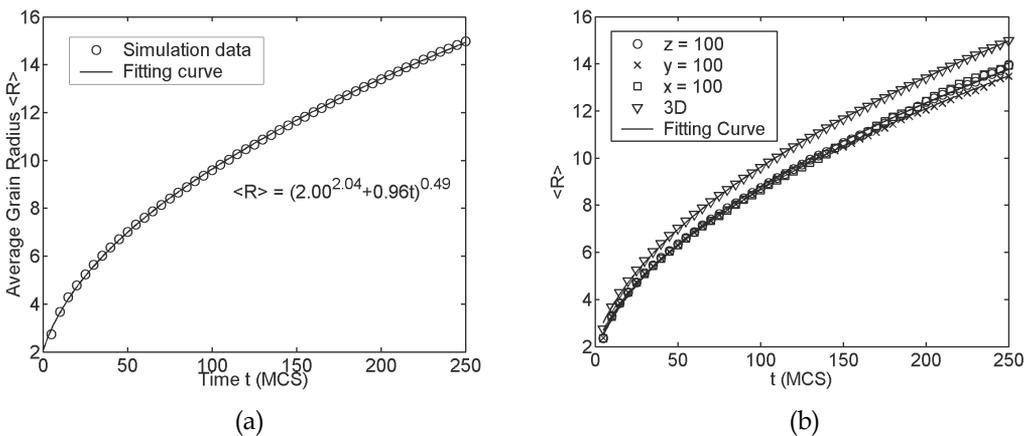


Fig. 16. Grain growth kinetics obtained with  $n \approx 0.5$  (a) for 3D domain and (b) for cross sections with  $x = 100$ ,  $y = 100$  and  $z = 100$ , respectively

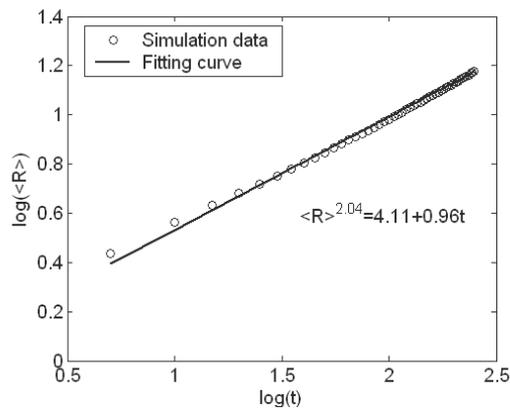


Fig. 17. Log-log plot of average grain radius  $\langle R \rangle$  vs. time  $t$  for 3D domain

## 6. Integration of Monte Carlo method into multi-scale approaches

Thermo-mechanical processing is one of the near-net-shape metal-processing technologies where a combination of mechanical pressure and heat is applied concurrently such as to deform a metallic workpiece into a desired shape. Since the microstructure of the manufactured products determines their mechanical properties, the various thermo-mechanical procedures (e.g., forging, rolling, extrusion, etc.) are heavily optimized in order to achieve advantageous microstructures. In the past, trial-and-error methods were used in industrial practice for designing and optimizing these thermo-mechanical processes, which led to high manufacturing costs and long lead times. Neither one of those is tolerable anymore in today's highly competitive and truly global economy. With modern computer technologies, numerical process modeling that is capable of predicting fairly accurately the shape of the deformed parts as well as the strain, strain rate, stress and temperature distributions has become feasible and more prevalent. This success of numerical process modeling is largely attributable to the achieved quality of the applied constitutive laws (i.e., coupling of mechanical and thermal models) that describe the basic material flow under the influence of pressure and heat and to the development of efficient and accurate numerical techniques such as the Finite Element Method (FEM; Altan & Vazquez, 1997; Walters et al., 1997).

Recent efforts toward improving the accuracy of the numerical predictions are now increasingly focusing on the modeling of the microstructural phenomena that occur during the thermo-mechanical processing of metals as consequences of complex metallurgical events such as recovery, recrystallization, grain growth, phase transformations, precipitation and dissolution reactions, etc. These microstructural phenomena may occur dynamically during deformation processing or either meta-dynamically or statically during post-deformation cooling or heat treatment. The mechanisms and kinetics of these phenomena and the associated changes in size, morphology, distribution, volume fraction and composition of the constituent phases are strongly dictated by the macroscopic heat flow and material flow processes (Humphreys & Hatherly, 1995; Doherty et al., 1997). Due to the variety and complexity of these microstructural events, such modeling presents significant challenges. Despite their scientific appeal, atomic-level modeling techniques require prohibitively high levels of computational power, and thus, they may remain

infeasible at least for the near future. Therefore, two types of modeling approaches are commonly applied, namely microstructure vs. processing-parameter relationships that were obtained empirically through regression analyses of experimental data as well as mesoscopic models that employ the physical laws governing the microstructure evolution.

In empirical models for the temporal evolution of the microstructure of metals during thermo-mechanical processing, the relationships between microstructural features (such as grain size, texture, topology, morphology at the grain or subgrain level, dislocation density, misorientation at the substructure level) and processing parameters (such as tool and workpiece geometry, temperature, deformation rate, amount of deformation, interface friction, heat transfer conditions) are derived via regression analyses of experimental data (Grong & Shercliff, 2002).

Currently, this empirical methodology is widely applied in industrial practice, albeit with moderate success. However, the range of applicability and the accuracy of such predictions are rather limited due to the empirical nature of the microstructure models employed. It should be noted that the underlying physical mechanisms of the microstructure evolution is not disclosed by these empirical models. Therefore, the applicability of these empirical models is confined to within the boundaries where they were obtained, and hence, they do not offer any universal prediction capabilities. Furthermore, because they are usually of a simple form, they are generally not suitable for describing more complicated microstructural phenomena.

Due to these shortcomings of empirical models, microstructural modeling approaches at the mesoscopic scale such as the Cellular Automaton (Feppon & Hutchinson, 2002) and the Monte Carlo Potts Model have been developed. In these approaches, the continuous material structure is discretized into a lattice that typically comprises thousands of grains. Physical laws (such as the surface energy reduction law governing normal grain growth, the site-saturated nucleation law for recrystallization, etc.) are then invoked to model the temporal evolution of the lattice. While mesoscopic models have been applied successfully to various microstructural phenomena in thermal processing such as welding (Yang et al., 2000) and film growth (Mizuseki et al., 2002), they have not yet been employed to predict the microstructure for industrially relevant thermo-mechanical processes such as hot forging or hot extrusion.

As a result of the recent rapid progress in the development of the mesoscopic microstructure modeling approach, various researchers have proposed to model the microstructure evolution in thermo-mechanical processing by combining a mesoscopic plasticity Finite Element (FE) analysis (Sarma et al., 1998) and the MC method (Radhakrishnan et al., 1998) or other microstructural models (Raabe & Becker, 2000) at the mesoscopic level. From a theoretical standpoint, these efforts are very appealing and the corresponding prediction methodology (Miodownik et al., 1999; Beaudoin et al., 2002) is expected to be extendable to problems of practically relevant size when macroscopic plasticity FE models have been obtained. However, macroscopic plasticity FE models may still not be available in the foreseeable future. Therefore, a systematic methodology for microstructure prediction in thermo-mechanical processing based on models at multiple length scales is briefly summarized below (Yu & Esche, 2005).

In such a multi-scale simulation methodology, the simulation input includes the processing conditions, the macroscopic and mesoscopic material properties as well as the initial microstructure features. These input parameters are supplied to the various modules of the modeling system, which consists of continuum-based coupled thermo-mechanical models, a

multi-scale modeling interface and mesoscopic microstructure models. The final output of the simulation sequence consists of the resulting microstructure features. For instance, consider the modeling of the microstructure evolution during hot forging of a single-phase material followed by an annealing procedure. In a simulation system consisting of FEM and MC modules, the simulation input includes the processing conditions (tool and workpiece geometry, temperature, deformation speed, amount of deformation, interface friction, etc.), the macroscopic material parameters (e.g., Young's modulus, Poisson's ratio, yield and hardening characteristics, thermal properties, etc.), initial microstructure features (i.e., initial grain size) and mesoscopic material properties (i.e., grain boundary energy and mobility). The simulation system is composed of an FEM module for modeling the macroscopic mechanical deformation process, a module for computing the stored energy from the primary field variables (i.e., strain and stress) and an MC-based module for modeling the static recrystallization and grain growth during the annealing process. The FEM-based calculations of the deformed system configuration and of the stored energy can be performed for the entire workpiece, which results in complete distributions of all field variables. Then, in order to save computational resources, the MC-based recrystallization and grain growth simulations can be limited to only a few representative elements of the FEM mesh as shown in Figure 18. Obtaining simulation results over the entire domain of the workpiece would probably necessitate a massively parallel computing system and might not even be necessary for most industrial applications. A more detailed discussion of the three simulation modules involved in the multi-scale modeling approach (i.e., FEM, multi-scale modeling interface and MC-based microstructure models) was provided elsewhere (Yu & Esche, 2005).

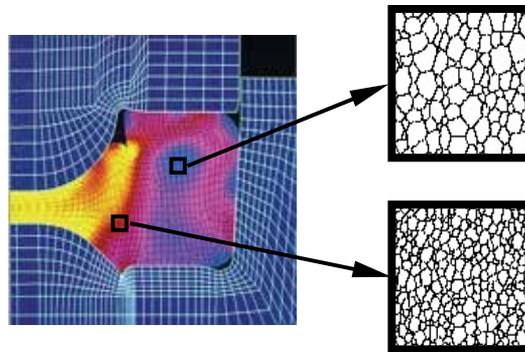


Fig. 18. Selection of representative zones for microstructure modeling<sup>1</sup>

The systematic multi-scale microstructure prediction methodology briefly described above represents one critically needed building block of a more comprehensive multi-scale modeling approach for material processing and material properties, which would be readily applicable to the industrial product design process. Such a comprehensive design system would comprise a continuum-based process model, a microstructure model and a material property model. Macro-scale process models are based on classical continuum mechanics. They predict the field variables as functions of the processing conditions and continuum material properties. These field variables in turn serve as the input to mesoscopic (or

<sup>1</sup> Figure adapted from website of Scientific Forming Technologies Corporation.

empirical) microstructure models for simulating the change in the microstructure features due to the material processing. Then, the resulting microstructure and field variables can be fed into material property models at either atomic or mesoscopic scales (or alternatively to empirical material property relationships). Finally, the material properties resulting from the material property models can either be compared with the given product requirements in order to optimize the processing conditions, or they can be inserted back into the process model such as to incorporate instantaneous influences of microstructural changes on the material properties used in the process model. In addition, the material property model can also provide mesoscopic material properties for the microstructure model.

## 7. Conclusion

In this chapter, the application of the MC method to two-dimensional simulations of the normal grain growth of isotropic single-phase materials was presented. After briefly reviewing some theoretical foundations on the self-similarity and the kinetics of grain growth, the conventional MC method was introduced. This was followed by a discussion of several modifications of the conventional MC algorithm that have been demonstrated to improve its accuracy and efficiency. Subsequently, the extension of the MC method to three-dimensional problems was described and its characteristics regarding grain growth kinetics and grain size distribution were analyzed. Finally, an approach for integrating the MC method into simulations of thermo-mechanical processes at multiple length scales was outlined.

## 8. References

- Altan, T. & Vazquez, V. (1997). Status of process simulation using 2D and 3D finite element method 'What is practical today? What can we expect in the future?'. *Journal of Materials Processing Technology*, Vol. 71, No. 1, pp. 49-63.
- Anderson, M. P., Srolovitz, D. J., Grest, G. S. & Sahni, P. S. (1984). Computer simulation of grain growth I – Kinetics. *Acta Metallurgica*, Vol. 32, No. 5, pp. 783-791.
- Anderson, M. P., Grest, G. S. & Srolovitz, D. J. (1989). Computer simulation of grain growth in three dimensions. *Philosophical Magazine*, Vol. B59, No. 3, pp. 293-329.
- Atkinson, H. V. (1988). Theories of normal grain growth in pure single phase systems. *Acta Metallurgica*, Vol. 36, No. 3, pp. 469-491.
- Beaudoin, A. J., Srinivasan, R. & Semiatin, S. L. (2002). Microstructure modeling and prediction during thermomechanical processing. *JOM*, Vol. 54, No. 1, pp. 25-29.
- Beenakker, C. W. J. (1988). Numerical simulation of a coarsening two-dimensional network. *Physical Review A*, Vol. 37, No. 5, pp. 1697-1702.
- Bolling, G. F. & Winegard, W. G. (1958). Grain growth in zone-refined lead. *Acta Metallurgica*, Vol. 6, No. 4, pp. 283-287.
- Bortz, A. B., Kalos, M. H. & Lebowitz, J. L. (1975). A new algorithm for Monte Carlo simulation of Ising spin systems. *Journal of Computational Physics*, Vol. 17, No. 1, pp. 10-18.
- Burke, J. E. & Turnbull, D. (1952). Recrystallization and grain growth. *Progress in Metal Physics*, Vol. 3, pp. 220-292.

- Chen, I.-W., Hassold, G. N. & Srolovitz, D. J. (1990). Computer simulation of final stage sintering II: Influence of initial pore size. *Journal of the American Ceramic Society* Vol. 73, No. 10, pp. 2865-2872.
- Doherty, R. D., Hughes, D. A., Humphreys, F. J., Jonas, J. J., Juul Jensen, D., Kassner, M. E., King, W. E., McNelley, T. R., McQueen, H. J. & Rollett, A. D. (1997). Current issues in recrystallization: a review. *Materials Science and Engineering A*, Vol. 238, No. 2, pp. 219-274.
- Feppon, J. M. & Hutchinson, W. B. (2002). On the growth of grains. *Acta Materialia*, Vol. 50, No. 13, pp. 3293-3300.
- Geiger, J., Roosz, A. & Barkoczy, P. (2001). Simulation of grain coarsening in two dimensions by cellular-automaton. *Acta Materialia*, Vol. 49, No. 4, pp. 623-629.
- Glazier, J. A., Gross, S. P. & Stavans, J. (1987). Dynamics of two dimensional soap froths. *Physical Review A*, Vol. 36, pp. 306-312.
- Grest, G. S., Srolovitz, D. J. & Anderson, M. P. (1985). Computer simulation of grain growth – IV. Anisotropic grain boundary energies. *Acta Metallurgica*, Vol. 33, No. 3, pp. 509-520.
- Grest, G. S., Anderson, M. P. & Srolovitz, D. J. (1988). Domain-growth kinetics for the Q-state Potts model in two and three dimensions. *Physical Review B*, Vol. 38, No. 7, pp. 4752-4759.
- Grong, O. & Shercliff, H. R. (2002). Microstructural modeling in metals processing. *Progress in Materials Science*, Vol. 47, No. 2, pp. 163-282.
- Hassold, G. N., Chen, I.-W. & Srolovitz, D. J. (1990). Computer simulation of final-stage sintering I: Model, kinetics, and microstructure. *Journal of the American Ceramic Society*, Vol. 73, No. 10, pp. 2857-2864.
- Hillert, M. (1965). On the theory of normal and abnormal grain growth. *Acta Metallurgica*, Vol. 13, No. 3, pp. 227-238.
- Holm, E. A., Glazier, J. A., Srolovitz, D. J. & Grest, G. S. (1991). Effects of lattice anisotropy and temperature on domain growth in the 2-dimensional Potts-model. *Physical Review A*, Vol. 43, No. 6, pp. 2662-2668.
- Holm, E. A. (1992). Modeling microstructural evolution in single-phase, composite and two-phase polycrystals. *Ph. D. thesis*, University of Michigan.
- Holm, E. A., Srolovitz, D. J. & Cahn, J. W. (1993). Microstructural evolution in two-dimensional two-phase polycrystals. *Acta Metallurgica and Materialia*, Vol. 41, pp. 1119-1136.
- Holm, E. A. & Battaile, C. C. (2001). The computer simulation of microstructural evolution. *JOM*, Vol. 53, No. 9, pp. 20-23.
- Holm, E. A., Hassold, G. N. & Miodownik, M. A. (2001). On misorientation distribution evolution during anisotropic grain growth. *Acta Materialia*, Vol. 49, No. 15, pp. 2981-2991.
- Humphreys, F. J. & Hatherly, M. (1996). *Recrystallization and Related Annealing Phenomena*. Pergamon.
- Ivasishin, O. M., Shevchenko, S. V. & Semiatin, S. L. (2004). Modeling of abnormal grain growth in textured materials. *Scripta Materialia*, Vol. 50, No. 9, pp. 1241-1245.
- Ivasishin, O. M., Shevchenko, S. V., Vasiliev, N. L. & Semiatin, S. L. (2003). 3D Monte Carlo simulation of texture-controlled grain growth. *Acta Materialia*, Vol. 51, No. 4, pp. 1019-1034.

- Janssens, K. G. F., Raabe, D., Kozeschnik, E., Miodownik, M. A. & Nestler, B. (2007). *Computational Materials Engineering – An Introduction to Microstructure Evolution*, Elsevier Academic Press, Oxford.
- Kad, B. K. & Hazzledine, P. M. (1997). Monte Carlo simulation of grain growth and Zener pinning. *Material Science and Engineering A*, Vol. 238, No. 1, pp. 70-77.
- Krill III, C. E. & Chen, L. Q. (2002). Computer simulation of 3-D grain growth using a phase-field model. *Acta Materialia*, Vol. 50, pp. 3057-3073.
- Kumar, S., Gunton, J. D. & Kaski, K. (1987). Dynamical scaling in the Q-state Potts model. *Physical Review B*, Vol. 35, pp. 8517-8522.
- Lee, H. N., Ryoo, H. S. & Hwang, S. K. (2000). Monte Carlo simulation of microstructure evolution based on grain boundary character distribution. *Material Science and Engineering A*, Vol. 281, No. 1, pp. 176-188.
- Liu, G., Yu, H., Song, X. & Qin, X. (2001). A new model of three-dimensional grain growth: theory and computer simulation of topology-dependency of individual grain growth rate. *Materials & Designs*, Vol. 22, No. 1, pp. 33-38.
- Louat, N. P. (1974). On the theory of normal grain growth. *Acta Metallurgica*, Vol. 22, No. 6, pp. 721-724.
- Matsubara, H. (1999). Computer simulations for the design of microstructural developments in ceramics. *Computational Materials Science*, Vol. 14, No. 1, pp. 125-128.
- Mehnert, K. & Klimanek, P. (1996). Monte Carlo simulation of grain growth in textured metals. *Scripta Materialia*, Vol. 35, No. 6, pp. 699-704.
- Mehnert, K. & Klimanek, P. (1997). Grain growth in materials with strong textures: 3D Monte Carlo simulations. *Computational Materials Science*, Vol. 9, No. 1-2, pp. 261-266.
- Messina, R., Soucail, M. & Kubin, L. (2001). Monte Carlo simulation of abnormal grain growth in two dimensions. *Material Science and Engineering A*, Vol. 308, No. 1-2, pp. 258-267.
- Miodownik, M. A. (2002). A review of microstructural computer models used to simulate grain growth and recrystallisation in aluminum alloys. *Journal of Light Metals*, Vol. 2, No. 3, pp. 125-135.
- Miodownik, M. A., Holm, E. A., Godfrey, A. W., Hughes, D. A. & LeSar, R. (1999). Multiscale modeling of recrystallization. *MRS Symposium Proceedings*, Vol. 538, pp. 157-162.
- Miodownik, M., Holm, E. A. & Hassold, G. N. (2000). Highly parallel computer simulations of particle pinning: Zener vindicated. *Scripta Materialia*, Vol. 42, No. 12, pp. 1173-1177.
- Mizuseki, H., Hongo, K., Kawazoe, Y. & Wille, L. T. (2002). Multiscale simulation of cluster growth and deposition processes by hybrid model based on direct simulation Monte Carlo method. *Computational Materials Science*, Vol. 24, No. 1, pp. 88-92.
- Mullins, W. W. (1956). Two-dimensional motion of idealized grain boundaries. *Journal of Applied Physics*, Vol. 27, No. 8, pp. 900-904.
- Mullins, W. W. (1986). The statistical self-similarity hypothesis in grain growth and particle coarsening. *Journal of Applied Physics*, Vol. 59, No. 4, pp. 1341-1349.
- Mullins, W. W. (1991). The statistical particle growth law in self-similar coarsening. *Acta Metallurgica and Materialia*, Vol. 39, No. 9, pp. 2081-2090.
- Mullins, W. W. (1998). Grain growth of uniform boundaries with scaling. *Acta Materialia*, Vol. 46, No. 17, pp. 6219-6226.

- Mullins, W. W. & Vinals, J. (1989). Self-similarity and growth kinetics driven by surface free energy reduction. *Acta Metallurgica*, Vol. 37, pp. 991-997.
- Nosonovsky, M., Zhang, X. & Esche, S. K. (2009). Scaling of Monte Carlo simulations of grain growth in metals. *Modelling and Simulation in Materials Science and Engineering*, Vol. 17, No. 2, 13 pp.
- Peczak, P. (1995). A Monte Carlo study of influence of deformation temperature on dynamic recrystallization. *Acta Metallurgica et Materialia*, Vol. 43, No. 3, pp. 1279-1291.
- Potts, R. B. (1952). Some generalized order-disorder transformations. *Proceedings of the Cambridge Philosophical Society*, Vol. 48, pp. 106-109.
- Raabe, D. (2000). Scaling Monte Carlo kinetics of the Potts model using rate theory. *Acta Materialia*, Vol. 48, No. 7, pp. 1617-1628.
- Raabe, D. & Becker, R. C. (2000). Coupling of a crystal plasticity finite-element model with a probabilistic cellular automaton for simulating primary static recrystallization in aluminum. *Modelling and Simulation in Materials Science and Engineering*, Vol. 8, No. 4, pp. 445-462.
- Raabe, D. (2002). Cellular Automata in materials science with particular reference to recrystallization simulation. *Annual Review of Materials Research*, Vol. 32, pp. 53-76.
- Radhakrishnan, B. & Zacharia, T. (1995). Simulation of curvature-driven grain growth by using a modified Monte Carlo algorithm. *Metallurgical and Materials Transactions A*, Vol. 26, pp. 167-180.
- Radhakrishnan, B., Sarma, G. B. & Zacharia, T. (1998). Modeling the kinetics and microstructural evolution during static recrystallization – Monte Carlo simulation of recrystallization. *Acta Materialia*, Vol. 46, No. 1, pp. 4415-4433.
- Rhines, F. N., Craig, K. R. & DeHoff, R. T. (1974). Mechanism of steady-state grain growth in aluminum. *Metallurgical Transactions*, Vol. 5, pp. 413-425.
- Rivier, N. (1983). On the structure of random tissues or froths, and their evolution. *Philosophical Magazine*, Vol. 47, No. 5, pp. L45-L49.
- Rollett, A. D., Srolovitz, D. J. & Anderson, M. P. (1989). Simulation and theory of abnormal grain growth – Variable grain boundary energies and mobilities. *Acta Metallurgica*, Vol. 37, No. 4, pp. 2127-1240.
- Rollett, A. D., Srolovitz, D. J., Anderson, M. P. & Doherty, R. D. (1992a). Computer simulation of recrystallization – III. Influence of a dispersion of fine particles. *Acta Metallurgica*, Vol. 40, No. 12, pp. 3475-3495.
- Rollett, A. D., Luton, M. J. & Srolovitz, D. J. (1992b). Computer simulation of dynamic recrystallization. *Acta Metallurgica et Materialia*, Vol. 40, No. 1, pp. 43-55.
- Rollett, A. D. & Mullins, W. W. (1996). On the growth of abnormal grains. *Scripta Metallurgica et Materialia*, Vol. 36, No. 9, pp. 975-980.
- Rollett, A. D. (1997). Overview of modeling and simulation of recrystallization. *Progress in Materials Science*, Vol. 42, No. 1-4, pg. 79-99.
- Rollett, A. D. & Raabe, D. (2001). A hybrid model for mesoscopic simulation of recrystallization. *Computational Materials Science*, Vol. 21, No. 1, pp. 69-78.
- Safran, S. A., Sahni, P. S. & Grest, G. S. (1983). Kinetics of ordering in two dimensions. I. Model systems. *Physical Review B*, Vol. 28, No. 5, pp. 2693-2704.
- Sahni, P. S., Srolovitz, D. J., Grest, G. S., Anderson, M. P. & Safran, S. A. (1983). Kinetics of ordering in two dimensions. II: Quenched systems. *Physical Review B*, Vol. 28, No. 5, pp. 2705-2716.

- Sarma, G. B., Radhakrishnan, B. & Zacharia, T. (1998). Finite element simulations of cold deformation at the mesoscale. *Computational Materials Science*, Vol. 12, No. 2, pp. 105-123.
- Sista, S. & DebRoy, T. (2001). Three dimensional Monte Carlo simulation of grain growth in zone refined iron. *Metallurgical and Materials Transactions B*, Vol. 32, pp. 1195-1201.
- Song, X. & Liu, G. (1998). A simple and efficient three-dimensional Monte Carlo simulation of grain growth. *Scripta Materialia*, Vol. 38, No. 11, pp. 1691-1696.
- Song, X. Y., Liu, G. Q. & He, Y. Z. (1998). Modified Monte Carlo method for grain growth simulation. *Progress in Natural Science*, Vol. 8, pp. 92-97.
- Song, X. & Rettenmayr, M. (2002). Modelling study on recrystallization, recovery and their temperature dependence in inhomogeneously deformed materials. *Materials Science and Engineering A*, Vol. 332, No. 1-2, pp. 153-160.
- Soucail, M., Messina, R., Cosnuau, A. & Kubin, L. P. (1999). Monte Carlo simulation of Zener pinning in two dimensions. *Material Science and Engineering A*, Vol. 271, No. 1, pp. 1-7.
- Srolovitz, D. J., Anderson, M. P., Sahni, P. S. & Grest, G. S. (1984a). Computer simulation of grain-growth: II. Grain size distribution, topology, and local dynamics. *Acta Metallurgica*, Vol. 32, No. 5, pp. 793-802.
- Srolovitz, D. J., Anderson, M. P., Grest, G. S. & Sahni, P. S. (1984b). Computer simulation of grain growth – III. Influence of a particle dispersion. *Acta Metallurgica*, Vol. 32, No. 9, pp. 1429-1438.
- Srolovitz, D. J., Grest, G. S. & Anderson, M. P. (1985). Computer simulation of grain growth – V. Abnormal grain growth. *Acta Metallurgica*, Vol. 33, No. 12, pp. 2233-2247.
- Srolovitz, D. J., Grest, G. S. & Anderson, M. P. (1986). Computer simulation of recrystallization – I. Homogeneous nucleation and growth. *Acta Metallurgica*, Vol. 34, No. 9, pp. 1833-1845.
- Srolovitz, D. J., Grest, G. S., Anderson, M. P. & Rollett, A. D. (1988). Computer simulation of recrystallization: II. Heterogenous nucleation and growth. *Acta Metallurgica*, Vol. 36, No. 8, pp. 2115-2128.
- Tikare, V., Holm, E. A., Fan, D. & Chen, L. Q. (1998). Comparison of phase-field and Potts models for coarsening processes. *Acta Materialia*, Vol. 47, No. 1, 363-371.
- Von Neumann, J. (1952). Discussion: grain shapes and other metallurgical applications of topology. In: *Metal Interfaces*, ASM, Cleveland.
- Walters, J., Kurtz, S., Wu, W.-T. & Tang, J. (1997). The 'state of the art' in cold forming simulation. *Journal of Materials Processing Technology*, Vol. 71, No. 1, pp. 64-70.
- Weiare, D. & Kermode, J. P. (1983). Computer simulation of a two-dimensional soap froth: I. Method and motivation. *Philosophical Magazine B*, Vol. 48, No. 3, pp. 245-259.
- Weiare, D. & Kermode, J. P. (1984). Computer simulation of a two-dimensional soap froth: II. Analysis of results. *Philosophical Magazine B*, Vol. 50, No. 3, pp. 379-388.
- Xiaoyan, S., Guoquan, L. & Nanju, G. (2000). Re-analysis on grain size distribution during normal grain growth based on Monte Carlo simulation. *Scripta Materialia*, Vol. 43, pp. 355-359.
- Yang, W., Chen, L.-Q. & Messing, G. L. (1995). Computer simulation of anisotropic grain growth. *Materials Science and Engineering A*, Vol. 195, pp. 179-187.
- Yang, Z., Sista, S., Elemer, J. W. & DebRoy, T. (2000). Three dimensional Monte Carlo simulation of grain growth during GTA welding of titanium. *Acta Materialia*, Vol. 48, pp. 4813-4825.

- Yu, Q. & Esche, S. K. (2003a). A Monte Carlo algorithm for single phase normal grain growth with improved accuracy and efficiency. *Computational Materials Science*, Vol. 27, No. 3, pp. 259-270.
- Yu, Q. & Esche, S. K. (2003b). A new perspective on the normal grain growth exponent obtained in two-dimensional Monte Carlo simulations. *Modelling and Simulation in Materials Science and Engineering*, Vol. 11, No. 6, pp. 859-862.
- Yu, Q. & Esche, S. K. (2003c). Three-dimensional grain growth modeling with a Monte Carlo algorithm. *Materials Letters*, Vol. 57, No. 30, pp. 4622-4626.
- Yu, Q. & Esche, S. K. (2004). Mesoscopic computer modeling of microstructure evolution within an object-oriented simulation framework. *International Journal of Computational Engineering Science*, Vol. 5, No. 3, pp. 451-469.
- Yu, Q., Wu, Y. & Esche, S. K. (2005). Modeling of grain growth characteristics in three-dimensional domains and two-dimensional cross sections. *Metallurgical and Materials Transactions A*, Vol. 36A, pp. 1661-1666.
- Yu, Q. & Esche, S. K. (2005). A Multi-scale approach for microstructure prediction in thermo-mechanical processing of metals. *Journal of Materials Processing Technology*, Vol. 169, pp. 493-502.

# Monte Carlo Simulations on Defects in Hard-Sphere Crystals Under Gravity

Atsushi Mori  
*The University of Tokushima*  
Japan

## 1. Introduction

In 1957 the crystalline phase transition in the hard-sphere system was discovered by a Monte Carlo simulation (Wood & Jacobson, 1957) and a molecular dynamics simulation (Alder & Wainwright, 1957). The results of their researches were surprising because the phase transition occurred despite the absence of attractive interparticle interaction. The crystalline phase transition in the hard-sphere system is sometimes referred to as the Alder transition or the Kirkood-Alder-Wainwright transition. Nowadays, the Alder transition can be interpreted as the competition between two entropic effects; if the configurational entropy overcomes the vibrational one, a disordered fluid phase appears as the stable phase and vice versa. The phase diagram was determined in 1968 by a Monte Carlo simulation (Hoover & Ree, 1968). It is temperature-independent and the phase transition from a disordered fluid phase to a crystalline phase of the face-centered cubic (fcc) structure occurs via a fluid-crystal coexistence region  $\phi_f < \phi < \phi_s$ . Here, the density is expressed by the volume fraction of the hard spheres,  $\phi \equiv (\pi/6)\sigma^3(N/V)$  with  $\sigma$  being the hard-sphere diameter,  $N$  the number of particles, and  $V$  the volume of the system. The Hoover and Ree's values ( $\phi_f = 0.494$  and  $\phi_s = 0.545$ ) have been revised by a direct crystal-fluid coexistence simulation (Davidchack & Laird, 1998) to be  $\phi_f = 0.491$  and  $\phi_s = 0.542$  as extending Mori et al.'s molecular dynamics simulation (Mori et al., 1995).

In 1960-70s colloidal crystallizations were extensively studied as the Alder transitions in reality. Indeed, an effective hard-sphere model successfully explained the colloidal crystal phase transition (Wadachi & Toda, 1972). The recent situation of studies on the colloidal crystal is different from that in those days; so-called hard-sphere suspensions are synthesized (Antl et al., 1986), which exhibit a hard-sphere nature in the crystal-fluid phase transition (Paulin & Ackerson, 1990; Phan et al., 1996; Pusey & van Megen, 1986; Underwood et al., 1994). There is another trend of studies of the colloidal crystals in recent days. Because in the colloidal crystals a periodic structure of dielectric constant with the periodicity of the same order of optical wavelength, the colloidal crystals can be used as photonic crystals. Ohtaka first pointed out this possibility (Ohtaka, 1979). Two 1987 papers triggered this trend (John, 1987; Yablonovitch, 1987). As compared to micro manufacturing technologies of fabricating the photonic crystals, the colloidal crystallization is of low cost in introducing equipment and less time consuming in the fabrication. One of shortcomings of the colloidal crystallization is that the colloidal crystals contain many crystal defects. From fundamental as well as application point of view,

the defect in the photonic crystal should be reduced. For example, the photonic band cannot be opened unless the defect is reduced.

In relation to the reduction of the crystal defects in the colloidal crystals, in 1997 Zhu et al. (Zhu et al., 1986) found an effect of gravity that reduces the stacking disorder in the hard-sphere colloidal crystals. They found that a random hexagonal close pack (rhcp) structure formed under micro gravity. On the other hand, the sediment is rhcp/fcc mixture under normal gravity (Pusey et al., 1989). The mechanism of reduction of the stacking disorder under gravity was so far unresolved until the present author and coworkers found a glide mechanism of disappearance of a stacking fault (Mori et al., 2007a). Viewing  $\langle 111 \rangle$  fcc is characterized by a stacking of ABCABC... sequence, where A, B, and C distinguish hexagonal planes on the basis of the position of the particles within the hexagonal plane. On the other hand, hexagonal close pack (hcp) structure is given by ABAB... stacking and rhcp by a random sequence of A, B, and C. The stacking disorder is the disorder in the sequence of A, B, and C. For example, an intrinsic stacking fault is given by a sequence such as ABABC...; here the third C plane has been removed from ABCABC... We note that even though the stacking is out of order, the particle density remains unchanged. In this respect, the varieties of stacking sequence are not affected by gravity. Thus, the mechanism of the reduction of the stacking disorder due to gravity was a long standing problem. To resolve this problem is the purpose of studies (Mori et al., 2006a;b; 2007a;b; 2009; Mori & Suzuki, 2010; Yanagiya et al., 2005) reviewed in section 4.1.

In a previous paper (Mori et al., 2007a) looking into the evolution of snapshots of Monte Carlo simulations of hard spheres (Mori et al., 2006b), in which transformation from a defective crystal into a less-defective crystal under gravity was observed, we found that a glide of a Shockley partial dislocation terminating an intrinsic stacking fault shrunk the stacking fault in fcc (001) stacking. The key is the fcc (001) stacking; in those simulations this stacking was forced due to a stress from a small periodic boundary simulation box. In contrast, in the colloidal crystallization a patterned bottom wall is sometimes used recently; the fcc (001) stacking is forced due to the stress from the pattern on the bottom. Use of a patterned bottom wall is called a colloidal epitaxy. In 1997 van Blaaderen et al. succeeded in the fcc (001) stacking using a fcc (001) pattern (van Blaaderen et al., 1997). The basic idea of the colloidal epitaxy is that the stacking sequence is unique in (001). The finding of a previous paper (Mori et al., 2007a) is that in the fcc (001) stacking, even if an intrinsic stacking fault running along oblique  $\{111\}$  plane is introduced, through the glide of a Shockley partial dislocation terminating the lower end of the stacking fault the stacking fault shrinks. In other words, their paper points out superiority of the colloidal epitaxy other than the unique stacking sequence. We note here that this glide mechanism is merely one of mechanism. The intrinsic stacking fault is mere one of metastable configurations. Moreover, we have already found a configuration which was succeeded into a newly grown crystal in the fluid phase in some simulation of the same condition (Mori et al., 2007b). An additional remark is that a coherent growth occurred in those simulations (Mori et al., 2006a). Complementarily to the simulations, we have given elastic energy calculations to understand the driving force of upward move of the Shockley partial dislocation (Mori et al., 2009; Mori & Suzuki, 2010).

We note again that in those simulations (Mori et al., 2006a;b; 2007a;b; Yanagiya et al., 2005), fcc (001) stacking was forced due to the stress from a small periodic boundary simulation box. This artifact has been resolved (Mori, in press). The same stress in those previous simulations can, in principle, be provided by the use of patterned substrate (the colloidal epitaxy). However, the system size cannot be systematically enlarged in those simulations

with the flat bottom wall. As already shown (Mori et al., 2006b), fcc {111} stacking occurs for a large lateral system size. In a recent paper (Mori, in press) the square pattern has been used. An advantage of the square pattern is that matching between the crystal grown and the substrate on the lattice line, not only on the lattice point, is possible (Lin et al., 2000). To resolve this shortcoming is the purpose of section 4.2.

The remainder of this chapter is organized as follows. In section 2 remarks on statistical mechanics of the hard-sphere system is described. Simulation method is reviewed in section 3. Results for flat bottom walls are reproduced and discussions for them are given in section 4.1 and some results and discussions for square patterned wall are presented in section 4.2. Conclusions and remarks are given in section 5.

## 2. Hard sphere system

The hard sphere system is comprised of the hard-sphere potential

$$V^{\text{HS}}(r_{ij}) = \begin{cases} \infty & r_{ij} \leq \sigma \\ 0 & r_{ij} > \sigma \end{cases}. \quad (1)$$

Here,  $r_{ij}$  is the interparticle separation between particles  $i$  and  $j$ . The total system energy is given by summing  $V_{ij} \equiv V(r_{ij})$  as

$$U = \sum_{(i,j)} V_{ij}, \quad (2)$$

where the summation is taken for all pairs. The configurational integral is defined as

$$\begin{aligned} Z &= \int \cdots \int dr_1 \cdots dr_N \exp[-U/k_{\text{B}}T], \\ &= \int \cdots \int dr_1 \cdots dr_N \prod_{(i,j)} \exp[-V_{ij}/k_{\text{B}}T], \end{aligned} \quad (3)$$

where  $k_{\text{B}}T$  is the temperature multiplied by Boltzmann's constant. In the integrand, by substituting  $V_{ij}^{\text{HS}}$  [Equation (1)] for  $V_{ij}$ , the quantity  $\exp[-V_{ij}/k_{\text{B}}T]$  takes either 0 or 1.

$$\exp[-V_{ij}/k_{\text{B}}T] = \begin{cases} 0 & r_{ij} \leq \sigma \\ 1 & r_{ij} > \sigma \end{cases}. \quad (4)$$

A special remark is that the commutation of the thermodynamic limit,  $N \rightarrow \infty$  with  $N/V$  kept a finite value, and the hard-sphere limit,  $V_{ij} \rightarrow V_{ij}^{\text{HS}}$ , is not guaranteed. After calculating the probability distribution using a continuum potential  $V_{ij}$  and then taking the hard-sphere limit is not appropriate. Thus, the probability distribution is no longer a continuum function. Monte Carlo simulations are, thus, performed on the basis of equation (4). That is, if any pairs of particle overlap after a Monte Carlo move, then the attempt configuration is rejected, and otherwise accepted. The interaction between a particle and vessel walls is treated in the same manner if the walls are hard bodies.

The present system is exerted to the gravitational field. Thus, the gravitational energy

$$U^{\text{g}} = \sum_{i=1}^N mgz_i, \quad (5)$$

is added to the hard-sphere interaction. Here,  $z_i$  is the  $z$ -coordinate of particle  $i$ . In equation (3),  $U$  is replaced with  $U^{\text{HS}} + U^g$ .

$$Z = \int \cdots \int dr_1 \cdots dr_N \prod_{ij} \exp \left[ -\frac{V_{ij}^{\text{HS}}}{k_B T} \right] \prod_k \exp \left[ -\frac{mg\sigma z_k}{k_B T \sigma} \right]. \quad (6)$$

Here, the dimensionless quantity  $mg\sigma/k_B T \equiv g^*$ , which plays a central role in the hard-sphere system under gravity, is referred to as the gravitational number or the gravitational constant. Accordingly, the attempt configuration after a Monte Carlo move of a particle  $k$  from  $\mathbf{r}_k \equiv (x_k, y_k, z_k)$  to  $\mathbf{r}_k + \Delta\mathbf{r} \equiv (x_k + \Delta x, y_k + \Delta y, z_k + \Delta z)$  is accepted with the probability

$$\exp[-g^* \Delta z^*], \quad (7)$$

in a usual manner, such as Metropolis' method, unless the overlap between particle  $k$  and any other ones occurred. Here,  $\Delta z^* \equiv \Delta z/\sigma$  is the change in  $z$  coordinate of particle  $k$  in unit of length of  $\sigma$  (hereafter, \* indicates this reduced unit).

### 3. Simulation method

#### 3.1 Stepwise control of the gravitational number

In a gravitational sedimentation the gravitational number  $g^*$  is controlled through the temperature  $T$ . In the previous work (Mori et al., 2006b) we proposed the stepwise control of  $g^*$  in order to avoid trapping of the system in a metastable configuration such as a polycrystalline state. Indeed, if  $g^*$  such as  $g^* \geq 0.9$  was turned on from the beginning, the system polycrystallized (Yanagiya et al., 2005). The basic idea of the stepwise  $g^*$  control is that in the simulated tempering (Lyubartev et al., 1992; Marinari & Parizi, 1992), unlike the simulated annealing (Kirkpatrick et al., 1983), no bias of lowering the temperature exists. We considered that if the system was relaxed with no such bias, the trapping into a metastable state such as a polycrystalline state might be avoided.

We note here that  $g^*$  can be controlled more effectively in centrifugation sedimentation of a colloidal dispersion. In the centrifugation method of the colloidal crystallization, such as done previously (Ackerson, 1999; Megens et al., 1997; Suzuki et al., 2007),  $g$  in  $g^* \equiv mg\sigma/k_B T$  can be directly controlled. Taking into account the fact that by the stepwise  $g^*$  control we could successfully avoid the trapping into a metastable state such as a polycrystalline state, a stepwise control of the centrifugation rotation velocity, or a more sophisticated control, must bring a effective control of the crystal defects in the colloidal crystals.

In this chapter, we reproduce results obtained by stepwise  $g^*$  control and will not presented results for simulations with sudden switch-on of gravity. For flat wall simulation (section 4.1) we kept  $g^*$  at a certain value for  $\Delta t = 2 \times 10^5$  Monte Carlo cycles and then increased by  $\Delta g^* = 0.1$ . Here, one Monte Carlo cycle is defined as it contains  $N$  Monte Carlo particle moves. That is, during one Monte Carlo cycle one Monte Carlo particle move is attempted per one particle on average. For square patterned wall simulations (section 4.2) we report results for  $\Delta t = 2 \times 10^5$  Monte Carlo cycles for  $N = 6656$  and  $\Delta t = 8 \times 10^5$  Monte Carlo cycles for  $N = 26624$ .  $\Delta g^* = 0.1$  for both system sizes.

#### 3.2 System size and configuration

In section 4.1 we reproduce results of Monte Carlo simulations of  $N = 1664$  hard spheres in a system with  $L_x^* = L_y^* = 6.27$  and  $L_z^* = 49.23$  (Mori et al., 2006b; 2007a). Flat hard walls were located at  $z=0$  and  $L_z$ .

After a random initial configuration at  $g^* = 0$  was prepared,  $g^*$  was increased as mentioned in section 3.1. We note that  $\phi = 0.45$  for this system was lower than  $\phi_f$ .  $L_z$  was enough large so that at  $g^*$  where the defect disappearance was observed a vacuum formed on the top. In horizontal ( $x$  and  $y$ ) directions the periodic boundary condition was imposed. In section 4.2 we report the results of Monte Carlo simulations of  $N = 6656$  and  $26624$  systems. The former is four times larger than  $N = 1664$ . Both  $L_x$  and  $L_y$  were doubled, i.e.,  $L_x^* = L_y^* = 12.55$ . The latter is four times larger than  $N = 6656$ ;  $L_x^* = L_y^* = 25.09$ . For both systems we set  $L_z^* = 200$  because volume of the vacuum region on the top of the system do not give affect to the crystal formed at the bottom if the vacuum region is enough large. There is an advantage in preparing a random initial configuration in a case of large  $L_z$ ; we should pay a spatial attention regarding surface ordering on the bottom and top walls and remnant of the crystalline order of the starting configuration if  $L_z$  is small. As for the flat wall case the periodic boundary condition was imposed in horizontal directions. A square patterned hard wall was put at  $z = 0$  and a flat hard wall at  $z = L_z$ . The square pattern is as illustrated in Figure 1. A grid made of square grooves with width  $0.70710678\sigma$  was formed. The diagonal distance of intersections of the longitudinal and transverse grooves was  $0.70710678\sigma \times \sqrt{2} = 0.9999999997\sigma$ . Thus, a hard sphere located on the lattice point of the bottom square lattice can fall into the intersection of the grooves, at most, by almost the half of the hard-sphere diameter; it means that the hard sphere cannot fall on to the bottom of the groove even if the groove depth is larger than  $0.5\sigma$ . The distance between edges of neighboring grooves was  $0.338\sigma$ .

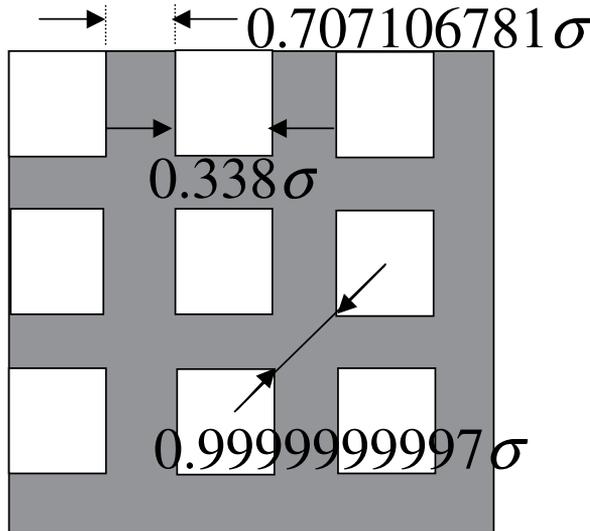


Fig. 1. Illustration of square pattern on the bottom wall.

## 4. Results and discussions

### 4.1 Flat wall case

We plot snapshots projected on  $xz$  plane for the flat bottom wall case in Figure 2. At  $g^* = 0.1$  the system was disordered except for the bottom crystalline layers [Figure 2(a)]. At  $g^* = 0.5$  a defective crystal was formed on the bottom and fluid phase above it [Figure 2(b)]. At  $g^* = 0.9, 1.3,$  and  $1.5$  we find a less-defective crystal on the bottom, a defective crystal above it, and

a fluid phase above the defective crystal [Figure 2(c-e)]. Comparing Figure 2 (c) and (d) we find that the boundary between the less-defective and defective crystals moved upward. Also the top of the defective crystal moved upward. On the other hand, comparing Figure 2 (d) and (e) we find that both the boundary between less-defective and defective crystals and the top of the defective crystal almost remained unchanged. Comparing Figure 2 (c)-(e) we find that the top of the fluid phase lowered with  $g^*$ . We can say that at  $g^* \sim 0.9$ , that is, when the gravitational energy  $mg\sigma$  was comparable to the thermal energy  $k_B T$ , the defect appearance occurred.

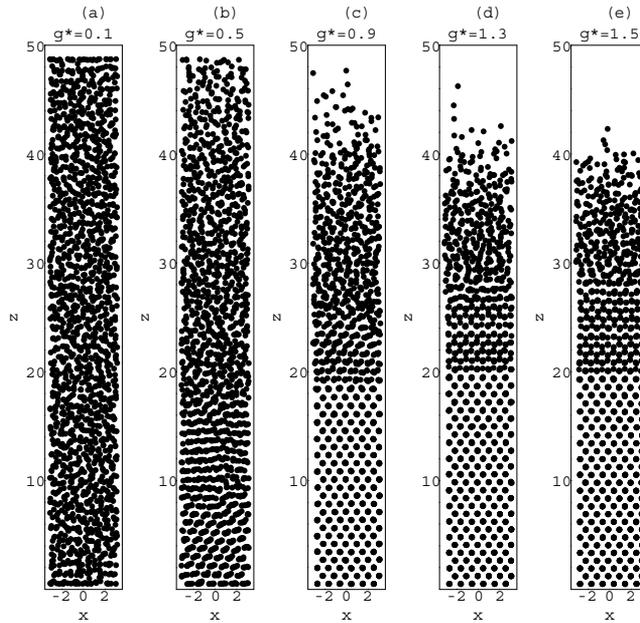


Fig. 2. Projections of snapshots for hard spheres in a system with flat wall; at (a)  $g^* = 0.1$ , (b) 0.5, (c) 0.9, (d) 1.3, and (e) 1.5. Snapshots at the end of duration while  $g^*$  was kept at the value indicated on the top of each figure were plotted. Reprinted with permission from THE JOURNAL OF CHEMICAL PHYSICS **124**, 174507 (2006). Copyright 2006, American Institute of Physics.

Before looking into a detail of the process of the defect disappearance, let us note that six crystalline layers exist along  $x$  and  $y$  axes, although  $L_x = L_y$  are four times the fcc lattice constant of the hard-sphere crystal at the crystal-fluid equilibrium. Moreover,  $[100]$  and  $[010]$  are no longer parallel to  $x$  and  $y$  axes, respectively. By a close look, it is found that  $x$  and  $y$  directions are, respectively, parallel to  $[110]$  and  $[\bar{1}\bar{1}0]$ . It means that the pressure at the bottom was higher than that at the crystal-fluid equilibrium. In a mechanical equilibrium

$$\frac{\partial P}{\partial z} = -mg\rho(z), \quad (8)$$

holds, where  $P(z)$  is the pressure at the altitude  $z$  and  $\rho(z)$  the coarse-scale number density at  $z$ . We can understand the higher pressure at the bottom according to this equation. If the relation between  $P$  and  $\rho$  (i.e., the equation of state) is known, we can solve equation (8) with

the condition

$$\int_{-L_x/2}^{L_x/2} \int_{-L_y/2}^{L_y/2} \int_0^{L_z} \rho dx dy dz = N, \tag{9}$$

such as done previously (Biben et al., 1993). Without doing so, we can understand the high pressure through an inequality of thermodynamic stability  $\partial P / \partial \rho > 0$ .

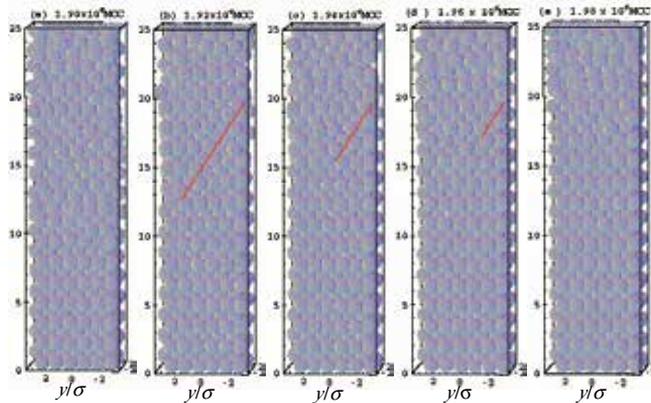


Fig. 3. 3D snapshots during defect disappearance occurred ( $g^* = 0.9$ ) for hard spheres in a system with flat wall; at (a) 1.99, (b) 1.92, (c) 1.94, (d) 1.96, and (e) 1.97  $\times 10^6$  Monte Carlo cycle. Reprinted with permission from Molecular Physics, Vol. 105, No. 10, 20 May 2007, 1377–1383. Copyright 2007, Taylor & Francis.

Let us look into the process of the defect disappearance. In Figure 3 evolution of the configuration during  $g^* = 0.9$  is shown in 3D. First of all, we notice that vertical stacking is basically two-fold, indicating the fcc (001) stacking; if the vertical stacking is fcc (111), it is three-fold, namely, ABC... This characteristic can be observed in Figure 2, too.

An intrinsic stacking fault is marked by a red line (an evidence that this defect is an intrinsic stacking fault will be given later). We see that the intrinsic stacking fault is shrinking in the course of the simulation. The altitude of the lower end of the stacking fault coincides with the  $z$  coordinate of the boundary of the less-defective and the defective crystals. We find that the transformation from defective crystal into the less-defective one observed in the two-dimensional snapshots (Figure 2) is accomplished by the shrinking of an intrinsic stacking fault. We note that if we used a different random number in a Monte Carlo simulation the position and the direction of the stacking fault were changed.

The evolution of the center of gravity during  $g^* = 0.9$  is plotted in Figure 4. The center of gravity moved downward overall as the simulation proceeded. Sinking of the center of gravity is understood by taking into account the particle deficiency of the dislocation core located at the lower end of the stacking fault. As the core moves upward the center of gravity sinks. We see plateaus during 1.84–1.9 and 1.96–2  $\times 10^6$  Monte Carlo cycles. This means that the system was trapped into metastable configurations during those durations. If the dislocation core goes up and enters into the fluid region, sinking of the center of the gravity finishes. The dislocation core in Figure 3 have not gone and not yet entered into the fluid region. So, if we continue the simulation longer than 2  $\times 10^5$  Monte Carlo cycles, further sinking of the center of gravity is expected. It is suggested that the defect disappearance in this case proceeded with temporal trapping by a metastable configuration, though we have not looked into the metastable configuration in the particle level yet.

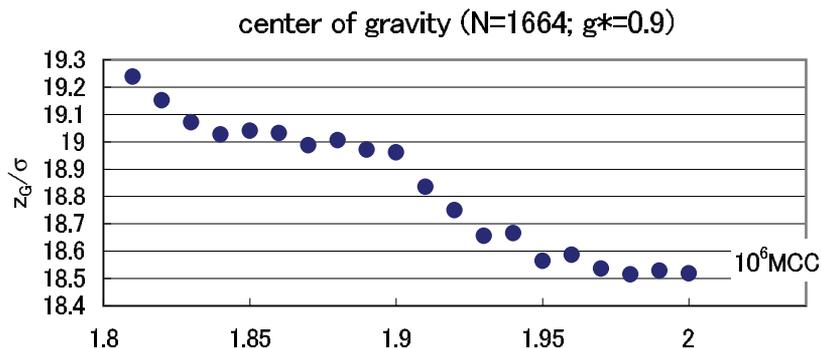


Fig. 4. Evolution of the center of gravity for hard spheres in a system with flat wall during  $g^* = 0.9$ . The centers of gravity are calculated at instant at every  $10^4$  Monte Carlo cycle. Reprinted with permission from Molecular Physics, Vol. 105, No. 10, 20 May 2007, 1377–1383. Copyright 2007, Taylor & Francis.

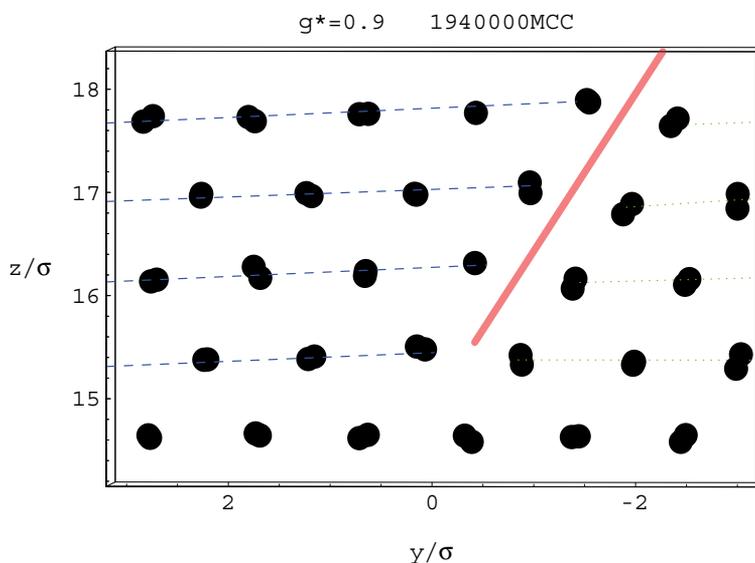


Fig. 5. Magnified snapshot of a defect for hard spheres in a system with flat wall (at  $1.94 \times 10^6$  Monte Carlo cycle). Reprinted with permission from Molecular Physics, Vol. 105, No. 10, 20 May 2007, 1377–1383. Copyright 2007, Taylor & Francis.

An evidence that the defect appearing in Figure 3 is an intrinsic stacking fault with a Shockley partial dislocation terminating its lower end can be given by looking into a magnified snapshot around the lower end of the defect. Figure 5 is a magnified snapshot around a lower end of the defect. The bottom layer includes no fault. The second and above layers include a fault, which is marked by a red line. We note that this fault is parallel to (111). The region up-left of this fault is shifted by  $\mathbf{b}^I = (1/6)[211] \equiv \mathbf{a}_1/3 + \mathbf{a}_2/6 + \mathbf{a}_2/6$ , where  $\mathbf{a}_1$ ,  $\mathbf{a}_2$ , and  $\mathbf{a}_2$  are the three lattice vectors.  $\mathbf{b}^I$  is the Burgers vector of a Shockley partial dislocation. Accordingly, at the lower end of the fault marked in Figure 5 a Shockley partial dislocation is formed. Because  $\mathbf{b}^I$  is the vector connecting a mid point of a upper triangle in

(111) triangular lattice (say, point B) to a mid point of the adjacent lower triangle in (111) lattice (say, point C), the stacking sequence around the fault is ABABC $\cdots$  or equivalent to this. In other words, removing the third C plane from ABCABC $\cdots$  is equivalent to shifting all planes right from the third C plane by  $b^I$ . In this way, the fault marked in Figure 5 is shown to be an intrinsic stacking fault. In a previous paper (Mori et al., 2007a), further, we observed shifts of magnitude of, respectively,  $a/2\sqrt{2}$ ,  $a/6\sqrt{2}$ , and  $a/6$  along [110],  $[\bar{1}\bar{1}0]$ , and [001]. Here,  $a$  is the fcc lattice constant. Readers may read a monograph (Hirth & Lothe, 1982) for learning the crystallography of defects.

#### 4.2 Squared patterned wall case

We have performed seven Monte Carlo simulations for  $N = 6656$  system and three for  $N = 26624$  with different random numbers. In two of these for  $N = 6656$  defect disappearance at  $g^*$  less than 0.9 was observed. Remember that the defect disappearance occurred during  $g^* = 0.9$  for the flat wall case (Mori et al., 2007a). In four of these for  $N = 6656$  the defect disappearance occurred at  $g^*$  greater than 0.9. For remainder one the defect disappearance was not appreciable. For  $N = 26624$  the defect disappearance occurred at  $g^*$  less than 0.9 for all three cases. In two of three the defect disappearance occurred during  $g^* = 0.5$  and in the remainder one during  $g^* = 0.7$ . The results below are essentially the same as a recent paper (Mori, in press).

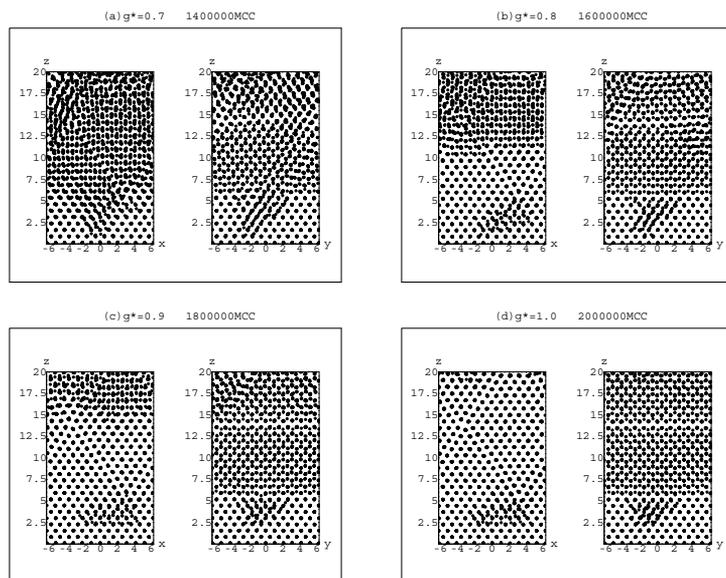


Fig. 6. Snapshot for 6656 hard spheres in a system with the square-patterned wall for a case that defect disappearance occurred at  $g^*$  lower than 0.9; at (a)  $g^* = 0.7$ , (b) 0.8, (c) 0.9, and (e) 1.0.

Figure 6 shows snapshots at  $g^* = 0.7-1.0$  for  $N = 6656$  system for a case that defect disappearance occurred at  $g^*$  less than 0.9. The random number for this simulation is different from that in a recent paper (Mori, in press). Throughout this section the random numbers for reported simulations are different from those in that paper. Though defects in lower portion remained, a defect in appearance expanded over the middle portion disappeared during  $g^* =$

1.0 again. In fcc (001) stacking, if a single stacking fault runs along one of  $\{111\}$  planes,  $[110]$  or  $[\bar{1}\bar{1}0]$  lattice line makes an array of two separated points in a projection on to  $(110)$  or  $(\bar{1}\bar{1}0)$ . And, on the other projection we can observe a fault directly. To understand Figure 6 we must take into account the fact that the  $x$  and  $y$  direction correspond to  $[110]$  and  $[\bar{1}\bar{1}0]$ , respectively (Mori et al., 2006b). In Figure 6 (a) splitting is observed in both  $xz$  and  $yz$  projections. Only the splitting in  $xz$  projection in portion  $7 < z^* < 11$  disappeared in Figure 6 (b). Also, the splitting in  $xz$  projection in portion  $z^* > 15$  in Figure 6 (c) disappeared in Figure 6 (d). We conjecture that two stacking faults such as along  $(111)$  and  $(\bar{1}\bar{1}1)$ , not  $(111)$  and  $(1\bar{1}\bar{1})$ , coexisted and then that along  $(111)$  shrunk. Splitting in both  $(110)$  and  $(\bar{1}\bar{1}0)$  is seen in a case that two stacking faults such as along  $(111)$  and  $(\bar{1}\bar{1}1)$  coexist. Making three-dimensional snapshot to observe intersections between  $(110)$  or  $(\bar{1}\bar{1}0)$  may give an direct answer to this conjecture. The surface structure of the 3D snapshots was, however, so complicated that we could not follow the edges of the stacking faults although we saw crossing two faults. We left these observations as a future research. Disappearance of only one of projections of lattice lines means the disappearance of the fault in corresponding direction and remaining of the fault in the other direction. There appear defects expanded in a lower-mid region, which remained stably throughout. We see an upper triangular shape at its right in  $xz$  projection and an lower triangular shape at its middle in  $yz$  projection. Upward and downward triangular shapes in projections imply the stacking fault tetrahedra. In (001) stacking, a tetrahedron surrounded by  $\{111\}$  makes upward and downward triangles in projections on to  $[110]$  and  $[\bar{1}\bar{1}0]$ , respectively. Identification of tetrahedra by observing the snapshot layer by layer traversing  $[001]$  is left for a future subject. The stacking fault tetrahedra are suggested to be sessile.

The evolutions of the center of gravity for  $N = 6656$  system during  $g^* = 0.8$  and  $1.0$  are plotted in Figure 7 for a case that defect disappearance occurred at  $g^*$  lower than  $0.9$ . During  $g^* = 0.8$  [Figure 7 (a)] the relaxation is of a single mode and has not reached to equilibrium yet. Figure 7 (a) is essentially the same as the corresponding figure in a recent paper (Mori, in press). Despite that the defect disappearance was proceeded, the system was not trapped into any metastable configuration. During  $g^* = 1.0$  [Figure 7 (b)], after a first relaxation, sinking of the center of gravity was slowed and then reached to equilibrium. The fluctuation after slowing (during  $1.87-1.9 \times 10^6$  Monte Carlo cycle) may be fluctuation around a metastable equilibrium. Sinking of the center of gravity in this duration might undergo a temporal stop as for the case of flat bottom wall. Those behaviors are observed in the corresponding figure in a recent paper (Mori, in press). Also, observation of metastable configuration in the particle level has not yet done.

Figure 8 shows snapshots at  $g^* = 0.7-1.0$  and  $1.3-1.4$  for  $N = 6656$  system for a case that defect disappearance occurred at  $g^*$  greater than  $0.9$ . We confirm no defect disappearance in Figure 8 (a)-(d). On the other hand, comparing Figure 8 (e) and (f) we find the defect disappearance in  $yz$  projection. The defect disappearance occurred during  $g^* = 1.4$  for this case is vary similar to that observed in Figure 6. What is suggested is essentially the same. The splitting of projection of lattice lines on  $yz$  direction disappeared during  $g^* = 1.4$ . This behavior is exactly the same as that reported in a recent paper (Mori, in press) except for the direction of the fault. Existence of a planner defect in the less defective portion  $z^* < 10$ , which seems as a line in projection, is observed in  $yz$  projection of Figure 8 (f) and  $xz$  projection of the corresponding figure in a recent paper (Mori, in press).

The evolutions of the center of gravity for  $N = 6656$  system during  $g^* = 1.0$  and  $1.4$  are plotted in Figure 9 for a case that defect disappearance occurred at  $g^*$  lower than  $0.9$ . That during  $g^* = 0.8$  essentially the same as Figure 7 (a) including the statistical error. We may regard the

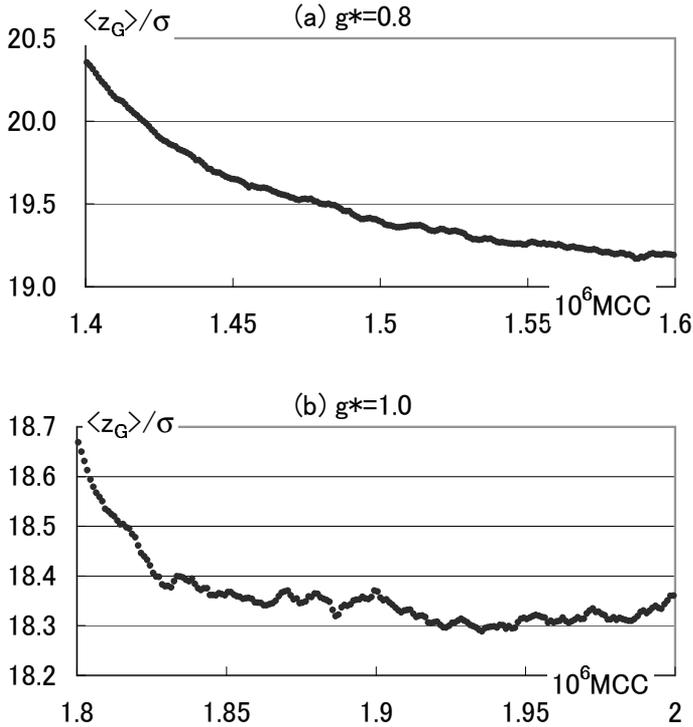


Fig. 7. Evolution of the center of gravity for 6656 hard spheres in a system with the square-patterned wall for a case that defect disappearance occurred at  $g^*$  lower than 0.9; during (a)  $g^* = 0.8$  and (b) 1.0. Running block average over  $10^3$  Monte Carlo cycles at every  $10^3$  Monte Carlo cycle is taken. Statistical errors are within  $0.011\sigma$  for (a) and  $0.008\sigma$  for (b).

sinking of the center of gravity during  $g^* = 1.0$  [Figure 9 (a)] to be of a single relaxation mode and a fluctuation around equilibrium as in a recent paper (Mori, in press). Unlike a recent paper (Mori, in press), the sinking during  $1.9\text{-}1.93 \times 10^6$  Monte Carlo cycle may be regarded as splitting a metastable state before this duration and an equilibrium state after that. However, as compared to Figure 9 (b) and the corresponding figure in a recent paper (Mori, in press), the multiple relaxation manner is not significant. Thus, the sinking of the center of the gravity during  $g^* = 1.0$  is of a single relaxation mode or of two stage manner with weak activation barrier between two stages. We can regard the sinking of the center of gravity in Figure 9 (b) to be a two stage manner; the system stay in a metastable equilibrium state during  $2.63\text{-}2.66 \times 10^6$  Monte Carlo cycle and then relaxes to an equilibrium state after  $2.75 \times 10^6$  Monte Carlo cycle. The two stage manner is more pronounced in a recent paper Mori (in press); the system stays in a metastable equilibrium state during  $2.63\text{-}2.71 \times 10^6$  Monte Carlo cycle and then relaxes to an equilibrium state after  $2.75 \times 10^6$  Monte Carlo cycle. An interesting thing is that despite shrinking of a "single" stacking fault is involved in Figure 6 (d) and Figure 8 (f), the sinking of the center of gravity in Figure 7 (b) and Figure 9 (b) is of a multiple manner. Of course, this is not surprising. Those behaviors are just similar to that in  $N = 1664$  small system with a flat bottom.

Figure 10 shows snapshots at  $g^* = 0.4\text{-}0.5$  and  $0.8\text{-}0.9$  for  $N = 26624$  system. Comparing  $xz$  projection of Figure 10 (a) and (b) we see that the splitting of the projection of lattice lines

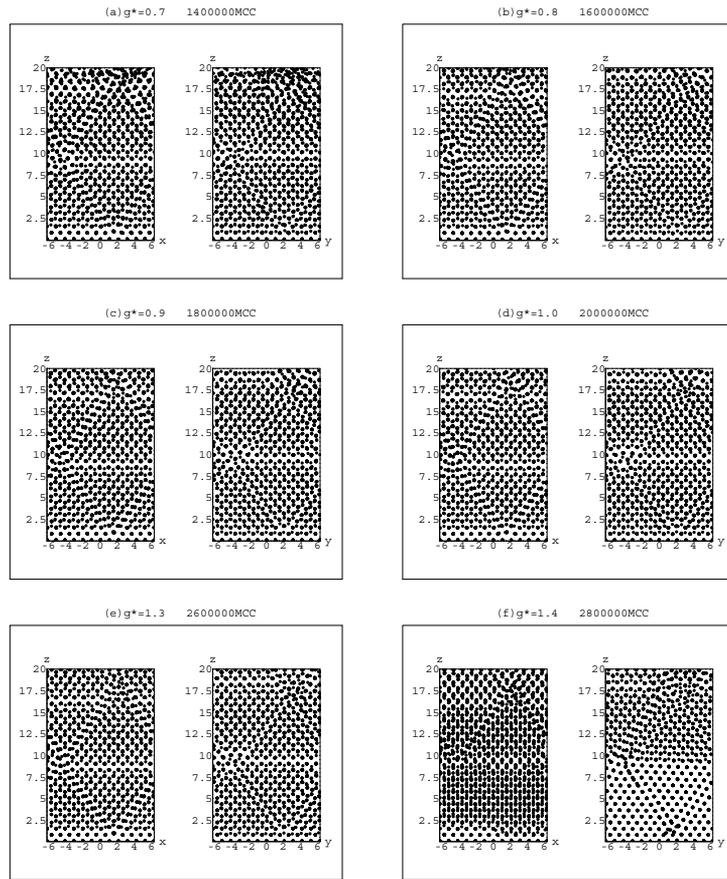


Fig. 8. Snapshot for 6656 hard spheres in a system with the square-patterned flat wall for a case that defect disappearance occurred at  $g^*$  higher than 0.9; at (a)  $g^* = 0.7$ , (b) 0.8, (c) 0.9, (d) 1.0, (e) 1.3, and (f) 1.4.

disappeared in portion  $3 < z^* < 10$ . This indicates shrinking of one or a few stacking faults running along  $(1\bar{1}1)$  or  $(1\bar{1}\bar{1})$  as discussed already. Comparing  $yz$  projection of Figure 10 (c) and (d) we see that the splitting of the projection of lattice lines disappeared in portion  $2.5 < z^* < 8$ . The splitting of the projection of lattice lines in  $xz$  direction has already disappeared at  $g^* = 0.8$ . This means shrinking of one or a few stacking faults running along  $(1\bar{1}1)$  or  $(1\bar{1}\bar{1})$  occurred. What is notable is the formation of triangular shapes both in  $xz$  and  $yz$  projections in Figure 10 (d). The suggestion of the stacking fault tetrahedron is as already discussed.

The evolutions of the center of gravity for  $N = 26624$  system during  $g^* = 0.5$  and 0.9 are plotted in Figure 9. During  $g^* = 0.5$  [Figure 11 (a)] the relaxation is of a single mode and has not reached to equilibrium yet. Figure 11 (a) is essentially the same as the corresponding figure in a recent paper (Mori, in press). Despite that the defect disappearance was proceeded, the system was not trapped into any metastable configuration. The activation barrier for the glide of dislocations or the motion of defect toward disappearance may become lower or vanishing as the system size becomes larger. Figure 11 (b) is quite similar to the corresponding figure

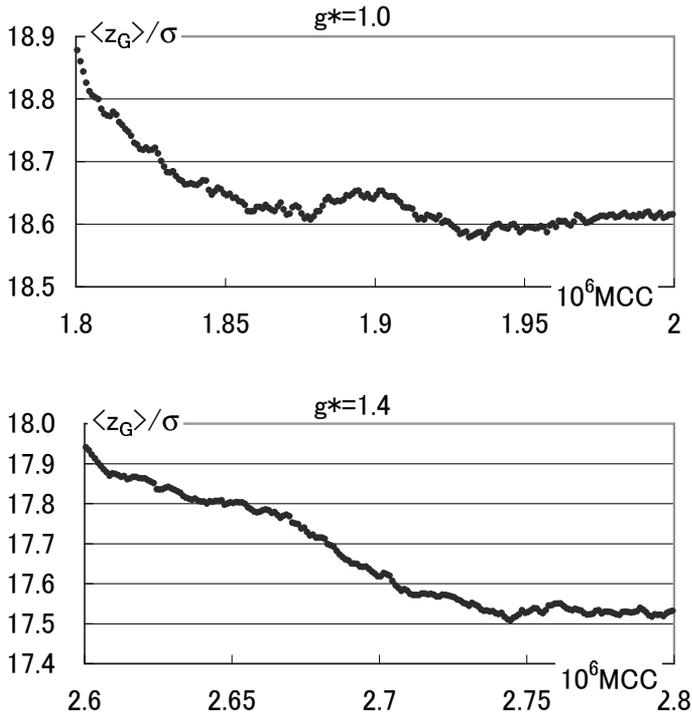


Fig. 9. Evolution of the center of gravity for 6656 hard spheres in a system with the square-patterned wall for a case that defect disappearance occurred at  $g^*$  higher than 0.9; during (a)  $g^* = 1.0$  and (b) 1.4. Running block average over  $10^3$  Monte Carlo cycles at every  $10^3$  Monte Carlo cycle is taken. Statistical errors are within  $0.010\sigma$  for (a) and  $0.007\sigma$  for (b).

in a recent paper (Mori, in press). It is of a single relaxation mode and fluctuation around equilibrium is observed.

## 5. Concluding remarks

We demonstrated the glide mechanism of a Shockley partial dislocation, which terminated an intrinsic stacking fault, for defect disappearance in hard-sphere colloidal crystal in fcc (001) stacking under gravity by Monte Carlo simulations (Mori et al., 2007a). This mechanism was seen at  $g^* \cong 0.9$  in a fcc (001) stacking crystal. Thus, we can say that we have pointed out a superiority of the colloidal epitaxy, which realizes the fcc (001) stacking.

However, the fcc (001) stacking in those simulations is forced by a stress from a small periodic boundary simulation box. That is, the driving force for the fcc (001) stacking was artificial. To resolve this shortcoming we have replaced the flat bottom wall with a square patterned one. By this mean, artificial driving force has been replaced with a realizable one. We have demonstrated the defect disappearance in those simulations.

In this chapter, we have concentrated on the defect disappearance and looked into the snapshot at relatively high  $g^*$ . Crystallization processes at low  $g^*$  have already been observed for the flat wall (Biben et al., 1994; Marechal & Dijkstra, 2007). Formation of a few crystalline layers at the bottom for the flat and square-patterned cases, details of which have been omitted

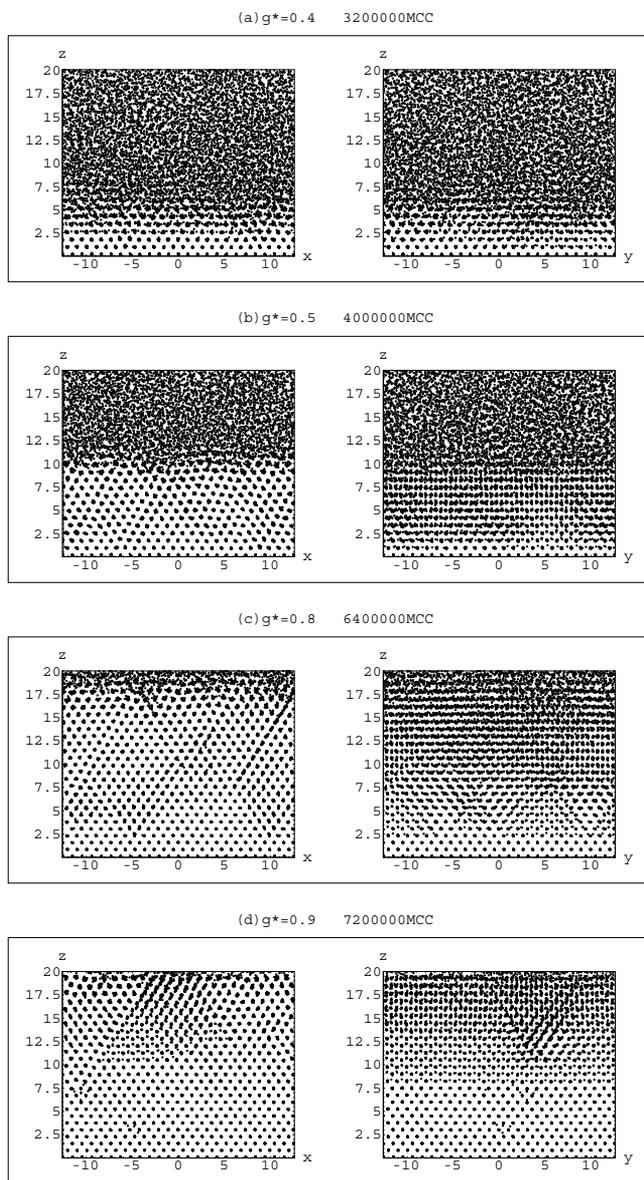


Fig. 10. Snapshot for 26624 hard spheres in a system with the square-patterned wall.

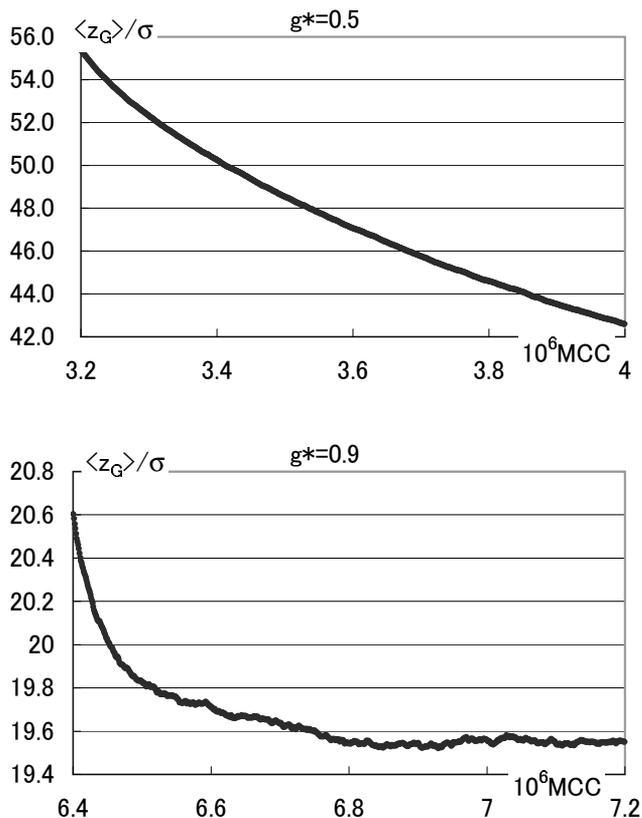


Fig. 11. Evolution of the center of gravity for 26624 hard spheres in a system with the square-patterned wall; during (a)  $g^* = 0.5$  and (b) 0.9. Running block average over  $10^3$  Monte Carlo cycles at every  $10^3$  Monte Carlo cycle is taken. Statistical errors are within  $0.0002\sigma$  for (a) and  $6 \times 10^5\sigma$  for (b).

in this chapter, is in agreement with the previous observation. More detailed analyses are in progress.

In simulations reported in this chapter, we adopted a conventional Monte Carlo method. Hence, the time corresponding to one Monte Carlo cycle varied. Indeed, acceptance ration varied depending on the density with a fixed maximum displacement of the Monte Carlo particle move. To perform kinetic Monte Carlo simulations to reproduce time evolution corresponding to real time is of interest. Also, molecular dynamics and Brownian dynamics simulations are planned.

## 6. References

- Ackerson, B. J.; Paulin, S. E.; Johnson, B.; van Meegen, W.; Underwood, S (1999). Crystallization by settling in suspension of hard spheres. *Physics Review E*, Vol. 59, No. 6, (June, 1999) 6903-6913, ISSN 1539-3755 (print), 1550-2376 (online), 1538-4519 (CD-Rom)
- Alder, B. J.; Wainwright, T. E. (1957). Phase Transition for a Hard Sphere System. *The Journal of Chemical Physics*, Vol. 27, No. 5, (November, 1957) 1208-1209, ISSN 0021-9606 (print), 1089-7690 (online)

- Antl, L.; Goodwin, J. W.; Hill, R. D.; Ottewill, R. H.; Owens, S. M.; Papworth, S. P.; Waters, J. A. (1986). The preparation of poly(methyl methacrylate) latices in non-aqueous media. *Colloids and Surfaces*, Vol. 17, No. 1, (January, 1986) 67-78, ISSN 0166-6622
- Biben, T.; Hansen, J.-P.; Barrat, J.-L. (1993). Density profiles of concentrated colloidal suspensions in sedimentation equilibrium. *The Journal of Chemical Physics*, Vol. 98, No. 9, (May, 1993) 7330-7344, ISSN 0021-9606 (print), 1089-7690 (online)
- Biben, T.; Ohnesorge, R.; Löwen, H. (1994). Crystallization in Sedimentation Profiles of Hard Spheres. *Europhysics Letters*, Vol. 28, No. 9, (December, 1994) 665-670, ISSN 0295-5075 (print), 1286-4854 (online)
- van Blaaderen, A.; Ruel, R.; Wiltzius, R. (1997). Template-directed colloidal crystallization. *Nature*, Vol. 385, No. 6614, (January, 1997) 321-324, ISSN 0028-0836 (print), 1476-4687 (online)
- Davidchack, R.; Laird, B. B. (1998). Simulation of the hard-sphere crystal-melt interface. *The Journal of Chemical Physics*, Vol. 108, No. 22, (June, 1998) 9452-9462, ISSN 0021-9606 (print), 1089-7690 (online)
- Hirth, J. H. & Loth, J. (1982). *Theory of Dislocations*, 2nd ed., Wiley, ISBN 0-89464-617-6, New York.
- Hoover, W. G.; Ree, F. H. (1968). Melting Transition and Communal Entropy for Hard Spheres. *The Journal of Chemical Physics*, Vol. 49, No. 8, (October, 1968) 3609-3617, ISSN 0021-9606 (print), 1089-7690 (online)
- John, S. (1987). Strong localization of photons in certain disordered dielectric superlattices. *Physical Review Letters*, Vol. 58, No. 23, (June, 1987) 2486-2489, ISSN 0031-9007 (print), 1079-7114 (online), 1092-0145 (CD-Rom)
- Kirkpatrick, S.; Gelatt, C. D.; Vecchi, P. (1983). Optimization by Simulated Annealing. *Science*, Vol. 220, No. 4598, (May, 1983) 671-680, ISSN 0036-8075 (print), 1095-9203 (online)
- Lin, K.-h.; Crocker, J. C.; Parasad, V.; Shofield, A.; Weitz, D. A.; Lubensky, T. C.; Yodh, Y. G. (2000). Entropically Driven Colloidal Crystallization on Patterned Surfaces. *Physical Review Letters*, Vol. 85, No. 8, (August, 2000) 1770-1773, ISSN 0031-9007 (print), 1079-7114 (online), 1092-0145 (CD-Rom)
- Lyubartev, A. P.; Martsinovski, A. A.; Shevkunov, S. V.; Vorontsov-Velyaminov, P. N. (1992). New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles. *The Journal of Chemical Physics*, Vol. 96, No. 3, (February, 1992) 1776-1783, ISSN 0021-9606 (print), 1089-7690 (online)
- Marechal, M.; Dijkstra, M. (2007). Crystallization of colloidal hard spheres under gravity. *Physical Review E*, Vol. 75, No. 6, (June, 2007) 061404-1-061404-8, ISSN 1539-3755 (print), 1550-2376 (online), 1538-4519 (CD-Rom)
- Marinari, E.; Parisi, G. (1992). Simulated Tempering: A New Monte Carlo Scheme. *Europhysics Letters*, Vol. 19, No. 6, (July, 1992) 451-458, ISSN 0295-5075 (print), 1286-4854 (online)
- Megens, M.; van Kats, C. M.; Bösecke, P.; Vos, W. L. (1997). Synchrotron small-angle X-ray scattering of colloids and photonic colloidal crystal. *Journal of Applied Crystallography*, Vol. 30, No. 5-2, (October, 1992) 637-641, ISSN 0021-8891 (print), ISSN 1600-5767 (online)
- Mori, A.; Manabe, R.; Nishioka, K. (1995). Construction and investigation of a hard-sphere crystal-melt interface by a molecular dynamics simulation. *Physical Review E*, Vol. 51, No. 5, (May, 1995) R3831-R3833, ISSN 1539-3755 (print), 1550-2376 (online), 1538-4519 (CD-Rom)

- Mori, A.; Yanagiya, S.-i.; Suzuki, Y.; Sawada, T.; Ito, K. (2006). Crystal structure of hard spheres under gravity by Monte Carlo simulation. *Science and Technology of Advanced Materials*, Vol. 7, No. 3, (April, 2006) 296-302, ISSN 1468-6996
- Mori, A.; Yanagiya, S.-i.; Suzuki, Y.; Sawada, T.; Ito, K. (2006). Monte Carlo simulation of crystal-fluid coexistence states in the hard-sphere system under gravity with stepwise control. *The Journal of Chemical Physics*, Vol. 124, No. 17, (May, 2006) 174507-1-174507-10, ISSN 0021-9606 (print), 1089-7690 (online)
- Mori, A.; Suzuki, Y.; Yanagiya, S.-i.; Sawada, T.; Ito, K. (2007). Shrinking stacking fault through glide of the Shockley partial dislocation in hard-sphere crystal under gravity. *Molecular Physics*, Vol. 105, No. 10, (May, 2007) 1377-1383, ISSN 0026-8976 (print), 1362-3028 (online); errata *ibid.* Vol. 106, No. 1, (January, 2008) 187-187
- Mori, A.; Suzuki, Y.; Yanagiya, S.-i. (2007). Succession of stacking disorder in hard-sphere crystal under gravity by Monte Carlo simulation. *Fluid Phase Equilibria*, Vol. 257, No. 2, (August, 2007) 131-138, ISSN 0378-3812
- Mori, A.; Suzuki, Y.; Matsuo, S. (2009). Disappearance of a Stacking Fault in Hard-Sphere Crystals under Gravity. *Progress of Theoretical Physics Supplement*, Vol. 178, (April, 2009) 33-40, ISSN 0375-9687 (print), 1347-0481 (online)
- Mori, A.; Suzuki, Y. (2010). Interplay between elastic fields due to gravity and a partial dislocation for a hard-sphere crystal coherently grown under gravity: driving force for defect disappearance. *Molecular Physics*, Vol. 108, No. 13, (July, 2010) 1731-1738, ISSN 0026-8976 (print), 1362-3028 (online)
- Mori, A. (in press). Monte Carlo simulation of growth of hard-sphere crystals on a square pattern. *Journal of Crystal Growth*, in press, ISSN 0022-0248
- Ohtaka, K. (1979). Energy band of photons and low-energy photon diffraction. *Physical Review B*, Vol. 19, No. 10, (May, 1979) 5057-5067, 1098-0121 (print), 1550-235X (online), 1538-4489 (CD-Rom)
- Paulin, S. E.; Ackerson, B. J. (1990). Observation of a phase transition in the sedimentation velocity of hard spheres. *Physical Review Letters*, Vol. 64, No. 22, (May, 1990) 2663-2666, ISSN 0031-9007 (print), 1079-7114 (online), 1092-0145 (CD-Rom); errata *ibid.*, Vol. 65, No. 5, (July, 1990) 668-668
- Phan, S. E.; Russel, W. B.; Cheng, Z.; Zhu, J.; Chaikin, P. M.; Dunsmur, J. H.; Ottewill, R. H. (1996). Phase transition, equation of state, and limiting shear viscosities of hard sphere dispersions. *Physical Review E*, Vol. 54, No. 6, (December, 1996) 6633-6645, ISSN 1539-3755 (print), 1550-2376 (online), 1538-4519 (CD-Rom)
- Pusey, P. N.; van Megen, W. (1986). Phase behaviour of concentrated suspensions of nearly hard colloidal spheres. *Nature*, Vol. 320, No. 6060, (March, 1986) 340-342, ISSN 0028-0836 (print), 1476-4687 (online)
- Pusey, P. N.; van Megen, W.; Bartlett, R.; Ackerson, B. J.; Parity, J. G.; Underwood, S. M. (1989). Structure of crystals of hard colloidal spheres. *Physical Review Letters*, Vol. 63, No. 25, (December, 1989) 2753-2756, ISSN 0031-9007 (print), 1079-7114 (online), 1092-0145 (CD-Rom)
- Suzuki, Y.; Sawada, T.; Mori, A.; Tamura, K. (1989). Colloidal Crystallization by Centrifugation. *Kobunsh Ronbumshu*, Vol. 64, No. 3, (March, 2007) 161-165 [in Japanese], ISSN 1881-5685 (print), 0386-2186 (online)
- Underwood, S. M.; Taylor, J. R.; van Megen, W. (1994). Sterically Stabilized Colloidal Particles as Model Hard Spheres. *Langmuir*, Vol. 10, No. 10, (October, 1994) 3550-3554, IPrint Edition ISSN 0743-7463, Web Edition ISSN 1520-5827

- Wadachi, M.; Toda, M. (1972). An Evidence for the Existence of Kirkwood-Alder Transition. *Journal of the Physical Society of Japan*, Vol. 32, No. 4, (April, 1972) 1147-1147, ISSN 0031-9015 (print), 1347-4073 (online)
- Wood, W. W.; Jacobson, J. D. (1957). Preliminary Results from a Recalculation of the Monte Carlo Equation of State of Hard Spheres. *The Journal of Chemical Physics*, Vol. 27, No. 5, (November, 1957) 1207-1208, ISSN 0021-9606 (print), 1089-7690 (online)
- Yablonovitch, E. (1987). Inhibited Spontaneous Emission in Solid-State Physics and Electronics. *Physical Review Letters*, Vol. 58, No. 20, (May, 1987) 2059-2062, ISSN 0031-9007 (print), 1079-7114 (online), 1092-0145 (CD-Rom)
- Yanagiya, S.-i.; Mori, A.; Suzuki, Y.; Miyoshi, Y.; Kasuga, M, Sawada, T.; Ito, K.; Inoue, T. (2005). Enhancement of Crystallization of Hard Spheres by Gravity: Monte Carlo Simulation. *Japanese Journal of Applied Physics*, Part. 1, Vol. 44, No. 7A, (July, 2005) 5113-5116, ISSN 0021-4922 (print), 1347-4065 (online)
- Zhu, J.; Li, M.; Rogers, R.; Meyer, W; Ottewill, R, STS-73 Space Shuttle Crew; Russel, W.; Chaikin, P. M. (1997). Crystallization of hard-sphere colloids in microgravity. *Nature*, Vol. 387, No. 6636, (June, 1997) 883-885, ISSN 0028-0836 (print), 1476-4687 (online)

# Atomistic Monte Carlo Simulations in Steelmaking: High Temperature Carburization and Decarburization of Molten Steel

R. Khanna, R. Mahjoub and V. Sahajwalla  
*Centre for Sustainable Materials Research and Technology,  
School of Materials Science and Engineering, University of New South Wales,  
Sydney, NSW2052,  
Australia*

## 1. Introduction

Carbon is one of the most important alloying elements in steel; a number of different properties have been attributed to strong interactions between interstitial carbon atoms and defects such as vacancies, dislocations and grain boundaries. The dissolution of carbonaceous materials such as graphite, coal, coke, chars etc. into molten iron is a key step in a range of ironmaking processes (Keogh et al., 1991; Cusack et al., 1991). While there have been some studies on the dissolution behavior of cokes and coal (Orsten & Oeters, 1988; Wright & Taylor, 1993; Mourao et al., 1993; Sahajwalla et al., 1994), most of the investigations have focused their attention on the dissolution behavior of graphite (Coller et al., 1994; Kosaka & Minowa, 1968; Ericsson & Melberg, 1981; Grigoryan & Karshin, 1972; Olivares, 1997). These studies both experimental and theoretical have provided a great deal of information about the reaction kinetics and various factors affecting the graphite dissolution rate, and have generally concluded that the dissolution of graphite is governed by mass transfer in the melt. This implies that the reactions at the interface are much faster than mass transfer and therefore do not control the dissolution kinetics. The dissolution behavior of coals, added either as lumps or fines, on the other hand is highly complex and involves volatile reactions and possible particle breakup. An understanding of the processes occurring at the carbon/melt interface is therefore of crucial importance; a fundamental understanding of the interfacial region is presently far from complete.

In addition to carburization from carbonaceous materials, liquid steel can pick up small quantities of carbon from carbon-based refractories, which are expended in the areas of hot metal treatment and transport, steelmaking vessels, steel secondary treatment as well as continuous casting (Sasai & Mizukami, 1995; Schei et al., 1988; Chen et al., 2000). In ironmaking processes including the Blast furnace and direct ironmaking technologies, significant quantities of carbon are used for the reduction of oxide iron ore as well as for the carburization of reduced metallic iron. High rates of carbon dissolution are generally required for high process efficiency. However during the final stages of refining of steel just prior to its casting into finished products, the composition of steel and relative concentrations of alloying elements are carefully controlled. The pick up of carbon from

refractories at this stage is highly undesirable both for steel as well as for the refractory, and needs to be minimised.

Another important aspect in steelmaking is on the role played by free surfaces and by surface-active elements (such as sulphur) on the decarburization of molten steel. During decarburization of continuously carbon-saturated liquid iron by  $\text{CO}_2$  between  $1280^\circ\text{C}$  and  $1600^\circ\text{C}$ , the presence of small amounts of sulphur (0.01 to 1 wt%) was found to significantly slow down the decarburization rate (Sain & Belton, 1978). Sulphur was assumed to cover the top monolayer of the liquid surface. Experiments showed a residual rate of decarburization, which was explained by assuming a fraction of sites, that could never be blocked by the surface active sulphur. The interfacial chemical rate was found to be independent of carbon levels for concentrations ranging between 1 wt% and saturation levels, and was of the first order with respect to the pressure of  $\text{CO}_2$  gas. The residual rate was also speculated as being either due to atomic misfit or the penetration of the adsorbed layer by the oxidant with a favourable orientation (Sain & Belton, 1976).

Due to immense technological importance of carburization and decarburization of molten iron for ironmaking & steelmaking operations, and with C and S playing key roles, our group at UNSW has been working in this field for the past twelve years. We have developed and optimised computer models of Fe-C-S and  $\text{Al}_2\text{O}_3$ -C/Fe systems for a fundamental atomic level understanding of high temperature processes taking place in molten iron and its interfaces with solid carbon and with alumina-carbon refractories. These atomic level investigations on a variety of steelmaking phenomena form the focus of this article.

This article is organised as follows. In Section 2, we present details of the atomic model of the Fe-C-S system, various interaction parameters and procedures used for their optimisation and validation for Monte Carlo computer simulations. Sections 3 to 5 respectively present atomistic simulations on the dissolution of carbon from graphite into molten iron, the depletion of carbon from  $\text{Al}_2\text{O}_3$ -C refractory by liquid steel and the role of free surfaces and surface active species on surface decarburization. All these sections also include specific details of computational algorithms, key results and comparisons with experimental results & novel findings. Section 6 will summarise key conclusions and ongoing/future research in this field.

## 2. Atomic model of Fe-C-S system and various interaction parameters

### 2.1 Fe-C System

In an attempt to understand atomic level interactions taking place at the graphite/Fe-C interface, we developed a theoretical model of the melt and the interfacial region designed specifically for lattice gas Monte Carlo simulation studies (Khanna & Sahajwalla, 1999). The main assumption of this model was regarding the structure of the liquid phase. While modeling of a liquid phase without access to direct experimental information, a generally good starting point is various crystalline structures observed for the alloy. There have been a few studies of the Fe-C solution phase where the atoms were assumed to occupy rigid lattice sites. In the interstitial model (Van Vlack, 1989), carbon atoms occupied octahedral interstitial sites with Fe atoms arranged on a regular fcc lattice; Monte Carlo simulations of the  $\alpha/\gamma$  phase boundary have been reported in literature using this model. A two sub-lattice model using defects and an associated solution model postulating molecular-like aggregates has also been used with varying degrees of success in binary systems containing liquid Fe (Guillermet et al., 1981; Jordan, 1979).

While considering the graphite/Fe-C melt interface in our investigation, atoms in molten iron were arranged on a rigid hexagonal lattice (space group: P6<sub>3</sub>/mmc). The sites in the melt were occupied by Fe and C atoms distributed randomly. This was a key assumption of this model. Although Fe and C atoms in a molten state are not expected to have a well defined structure, lattice sites need to be pre-defined for lattice gas Monte Carlo simulations. For ease of operation, atoms in Fe-C melt were placed on a rigid hexagonal lattice; there were no vacant sites and the contact surface between graphite and Fe-C melt was assumed to be smooth. A cubic structure of the melt would have caused unnecessary complications of boundary mismatch across the graphite/liquid iron interface. An attempt was made to account for the liquid nature of Fe-C through isotropic pair-wise short-range interaction parameters between the atoms; atomic interactions were however anisotropic in solid graphite.

The range of interaction was restricted to nearest neighbors (nn) in the basal plane and next nearest neighbors (nnn) along the c direction. The nn interaction, labeled as J<sub>1</sub>, accounted for the bonding in the basal plane. The interaction between the nearest neighbors along c direction, labeled as J<sub>2</sub>, accounted for the inter-layer bonding. The interactions were anisotropic for solid graphite with J<sub>2</sub> << J<sub>1</sub>. Isotropic interactions in the liquid Fe-C phase yielded J<sub>1</sub>=J<sub>2</sub>. While studying the ordering behavior of a disordered binary (AB) alloy (Mori et. al., 1963), interactions J<sub>1</sub> (& J<sub>2</sub>) can be replaced by a single ordering parameter: J = -(J<sub>AA</sub>+J<sub>BB</sub>-2J<sub>AB</sub>)/4. This simplification, however, is not possible in the present case and each interaction has to be taken into account explicitly.

Due to graphite being a solid at the temperatures of interest (up to 1600°C), both J<sub>1</sub> and J<sub>2</sub> were chosen to be attractive for C-C pairs; net bonding needs to be attractive to hold the atoms together. The parameter J<sub>1</sub> (and J<sub>2</sub>) was chosen to be repulsive for Fe-Fe pairs in the melt. The saturation concentration of C in Fe-C melts ranges from 4 to 6 wt% at a range of steelmaking temperatures (Chipman, 1972). Any additional C is known to precipitate out of the melt which implies a clustering of C atoms. From ordering energy considerations, the Fe-C interaction needs to be repulsive in nature for C-C clustering to take place. An attractive Fe-C interaction will lead to a disordered system representing a homogenous solution. While the sign of Fe-C interaction could be broadly determined by its ordering behavior, its magnitude depends on the enthalpy of mixing of Fe and C and also on the Gibbs energy of graphite (Lakaze & Sundman, 1991). The cohesive energy of graphite also determines the strength of C-C interaction. Other interaction parameters were represented in units of the nearest neighbor C-C interaction strength. Representing atoms as magnetic spins (S = +1 (up) for carbon and S = -1 (down) for iron), the Hamiltonian E of the system can be written as

$$E = -\sum_{i \neq j}^{nn} J_1(a - \beta) S_i^a S_j^b - \sum_{i \neq j}^{nnn} J_2(a - \beta) S_i^a S_j^b - H \sum_i S_i \quad (1)$$

where spin S<sub>i</sub><sup>α</sup> represents the type of atom (α) occupying the site i and J's are the various interaction parameters. Simulations were carried out using the following set of interaction parameters: J<sub>1</sub>(C-C) = J; J<sub>2</sub>(C-C) = γJ<sub>1</sub>(C-C); J<sub>1</sub>(Fe-Fe) = -J; J<sub>2</sub>(Fe-Fe) = J<sub>1</sub>(Fe-Fe); J<sub>1</sub>(Fe-C) = J<sub>2</sub>(Fe-C). γ was varied in the range 0.05-0.2 and J<sub>1</sub>(Fe-C) was varied in the range 0.4J to 0.7 J. The parameter J has units of energy and is assumed to be positive in magnitude.

Simulations were carried out in the grand canonical ensemble. A collection of N spins was placed on a hexagonal lattice; the relative concentration of up and down spins was controlled

by the carbon content in the melt. The magnetic field  $H$  in Eq. (1) was held constant and the relative concentration of spins was allowed to fluctuate. Starting from an initial configuration, a spin was chosen randomly and was flipped ( $S_i \rightarrow -S_i$ ). The energy difference  $\Delta E$  resulting from the spin flip was calculated. The flipping of spin was accepted for  $\Delta E \leq 0$ . For  $\Delta E > 0$ , the change may be accepted with a transition probability  $W$  (Binder et al., 1981)

$$W = \exp(-\Delta E / k_B T) / [1 + \exp(-\Delta E / k_B T)] \quad (2)$$

where  $k_B$  is the Boltzmann constant and  $T$  the temperature.  $W$  was compared to a random number  $\eta$  chosen uniformly between 0 and 1. The move was accepted for  $W > \eta$ ; otherwise old configuration was counted once more for averaging. This algorithm leads to a thermal equilibrium distribution in the limit when the number of generated states tend to infinity. In practice, fairly accurate results can be achieved with few thousand Monte Carlo steps per site. Lattice sizes for simulation ranged between  $24 \times 24 \times 24$  to  $30 \times 30 \times 30$ ; typically  $10^6$  to  $10^8$  MC steps were used in each simulation. A dimensionless parameter,  $T^*$  ( $= k_B T / J$ ) was used to represent reduced temperature.

A very well defined first order phase transition was observed between graphite on the carbon rich end and an ordered solution of Fe-C on the low carbon end. The transition was very sharp for  $J_1(\text{Fe-C}) = 0.6\text{J}$  and  $0.7\text{J}$  and was somewhat broad for  $J_1(\text{Fe-C}) = 0.4\text{J}$  and  $0.5\text{J}$ . A discontinuous behavior was observed in the plots of energy and order parameters. This phase transition was also investigated for the magnitude of  $\gamma$  ranging between 0.05 to 0.2; the transition became slightly broader with increasing  $\gamma$ . Apart from causing a general shift, the magnitudes of C-C and Fe-Fe interaction parameters did not show any new features. The simulated results compared very well with the experimental phase boundary (Chipman, 1972) and a correspondence was established between the experiment and theory. For example, an experimental study at  $1400^\circ\text{C}$  corresponded to a carbon solubility limit of 4.88 wt%. For  $J_1(\text{C-Fe}) = 0.5\text{J}$  and  $0.6\text{J}$ , this corresponded to  $T^* = 0.465$  and  $0.667$  respectively. The corresponding  $T^*$  values for  $1600^\circ\text{C}$  were 0.503 and 0.717 respectively. The magnitude of  $J_1(\text{Fe-C})$  could be chosen either as  $0.5\text{J}$  or  $0.6\text{J}$  without any loss of generality. The following sets of interaction parameters were used for dissolution studies:  $J_1(\text{C-C}) = J$ ,  $J_2(\text{C-C}) = \gamma J$ ;  $J_1(\text{Fe-Fe}) = J_2(\text{Fe-Fe}) = -J$ ;  $J_1(\text{C-Fe}) = J_2(\text{C-Fe}) = 0.5\text{J}$  and  $\gamma = 0.02$ .

## 2.1 Fe-C-S System

The effect of sulphur on the solubility of graphite in iron melts was investigated in the temperature range  $1400$ – $1600^\circ\text{C}$  based on the atomistic model of Fe-C system. It is well known that C and S atoms are strongly repulsive in the Fe-C-S system, and S atoms also repel each other (Ohtani & Nishizawa, 1986). The attractive bond between Fe and S is very strong and is more or less ionic in nature. This strong Fe-S bond is capable of distorting the electron distribution around the Fe atom also affect other bonds made by it (Kitchener et al., 1948). In the event of such a distortion taking place, the resulting bond energy will be a fraction  $(1-\epsilon)$  of the energy of the bond made in the absence of a Fe-S bond, with  $\epsilon$  ranging between 0 and 1.

Representing atoms as magnetic spins ( $S=+1$  for carbon,  $S=-1$  for iron and  $S=0$  for sulphur), the Hamiltonian  $E$  of the system in the Ising model can be written as

$$E = - \sum_{i \neq j}^{nn} [J_1(a - \beta) S_i^a S_j^b + K_{ij} R_1(a - \beta)] - \sum_{i \neq j}^{nnn} [J_2(a - \beta) S_i^a S_j^b + K_{ij} R_2(a - \beta)] - H \sum_i S_i \quad (3)$$

The constant  $K_{ij}$  has a value of 1 if either one or both sites  $i$  and  $j$  are occupied by sulphur and is zero otherwise. Values of  $R$  represent interactions of S with other atoms. The coefficient  $J_s$  and  $R_s$  have units of energy.  $H$  is the magnetic field.

Let  $J$  represent the magnitude of the nearest-neighbour C-C interaction strength. Simulations were carried out using the following set of interaction parameters for sulphur based interactions:  $R_1(S-S)=R_2(S-S)=-(0.1 \text{ to } 0.5)J$ ;  $R_1(S-C)=R_2(S-C)=-(0.1 \text{ to } 0.5)J$ ;  $R_1(Fe-S)=R_2(Fe-S)=(0.2 \text{ to } 1)J$ . The bonds made by Fe atoms which have at least one bond with a sulphur atom were modulated by a factor  $(1-\epsilon)$  where  $\epsilon$  takes on three values: 0.0, 0.5 and 1. Two values of  $\gamma$  (0.02 and 0.2) were used in these simulations. Detailed simulations on the system showed that the strength of the Fe-C interaction ( $J_1(Fe-C)=0.5J$  or  $0.6J$ ) did not have much effect on the linear trend of the decreasing carbon solubility with sulphur. It also did not affect the magnitude of the slope to a great extent. The relative strength of interlayer C-C bonding ( $\gamma$ ) was also not one of the crucial parameters in controlling the effect of sulphur on carbon solubility.

The strength of the strongly attractive Fe-S interaction appeared to be one of the important parameters. At 1400°C, and with no distortion of electron distribution around Fe atoms ( $\epsilon=0.0$ ), the simulated slopes  $m$  were closest to the experimental values for  $R_1(Fe-S)=0.5J$  with  $R_1(S-S)$  [and  $R_1(C-S)$ ] ranging from  $-0.1J$  to  $-0.5J$ .  $R_1(Fe-S)=0.2J$  gave lower values of the slope and  $R_1(Fe-S)=1.0J$  gave higher values. Overall slope values were slightly higher at 1600°C. With the Fe-S bond distorting the electron distribution around the Fe atom, either by an intermediate amount ( $\epsilon=0.5$ ) or by a large amount ( $\epsilon=1.0$ ), the slopes were no longer dependent on the actual magnitude of  $R_1(Fe-S)$  and  $R_1(S-S)$  [and  $R_1(C-S)$ ] and showed a rather flat range. The variation in slopes was within simulation error bars. The strength of the Fe-S interaction and its effect on the electron distribution around Fe atoms appeared to be one of the most important factors in determining the effect of sulphur on the solubility of graphite in Fe-C-S melts.

We extended this model further with the aim of optimizing various interaction parameters (Sahajwalla & Khanna, 2002) and to investigate the effect of electronic distortions around Fe and their relative importance in the molten state. Apart from a decrease in carbon solubility, the validated model needed to simultaneously satisfy other well-known features of this system. In the absence of carbon, the Fe-S system can be regarded as a solution of FeS in the iron melt. Even though this solution is non-ideal, it is completely miscible and forms a homogeneous phase. However the addition of small amounts of carbon cause Fe-C-S solution to separate into two immiscible layers, one rich in sulfur but low in carbon and other rich in carbon but low in sulphur (Morris & Buehl, 1950; Morris & Williams, 1949).

An atomic model, which can account for these key features of the Fe-C-S system, can be used to identify optimum strengths of various interaction parameters and form the basis for further studies on this system. As C and S atoms tended to displace each other to regions of high and low concentrations, it was assumed that this displacement was mediated by an Fe atom. Treating both C and S atoms on an equal footing and assuming electronic distortions around Fe play a significant role in this displacement process, two new parameters ( $\delta$ 's) were defined.  $\delta(Fe-C)$  represents the modification in the Fe-C interaction parameter, when the Fe atom has an additional bond with S. Similarly  $\delta(Fe-S)$  represents modification in the Fe-S interaction parameter, when the Fe atom has an additional bond with C. These parameters were varied over a wide range.  $\delta(Fe-S)$  ranged from  $-0.5$  to  $1.0$  with  $\delta(Fe-C)=1.0$ . It was expected that a locally repulsive Fe-S interaction may lead to a displacement of S from C's neighbourhood.

Simulation results on a homogenous Fe-C-S system showed that the liquid separated into two immiscible regions only for  $\delta(\text{Fe-S}) = 1.0$ . This indicates that distortion around Fe did not significantly affect Fe-S interaction strength and could be neglected. This separation was however most pronounced for  $\delta(\text{Fe-C}) = 1.5$ . Phase diagram simulations on carbon solubility also led to similar conclusions. Small values of  $\delta(\text{Fe-S})$  ( $-0.5$  and  $0.0$ ), which were found unsuitable in miscibility studies, also showed negligible effect of sulfur on carbon solubility. The parameter,  $\delta(\text{Fe-C}) = 0.5$ , was found to be completely unsuitable as it led to a slight increase in solubility rather than a decrease. Optimum parameters for this system, which simultaneously simulate the well known properties of Fe-C-S system were determined as:  $\epsilon(\text{Fe-Fe}) = 1.0$ ,  $\delta(\text{Fe-S}) = 1.0$ ,  $\delta(\text{Fe-C}) = 1.0$  and  $1.5$ . Simulation results on the Fe-C-S system clearly showed that distortions around Fe due to a strong Fe-S bond did not play a significant role in the molten state.

### 3. Dissolution of carbon into molten iron

Two identical blocks with hexagonal sites were placed in contact with each other (Fig. 1). All sites on the first block representing solid graphite were occupied by carbon. Sites on the second block representing Fe-C system were occupied by Fe, C atoms distributed randomly. The relative concentrations of Fe, C and S atoms were governed by the carbon and sulphur content of the melt under consideration.

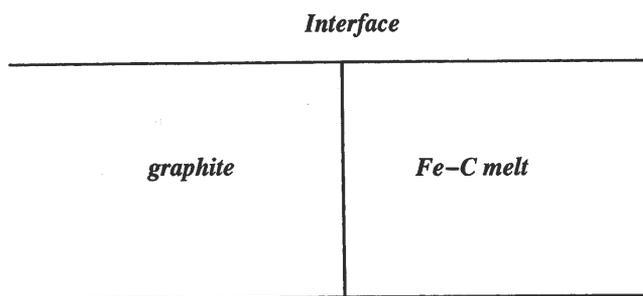


Fig. 1. A schematic representation of the simulation arrangement

The dissolution rate of graphite can be defined in terms of the number of carbon atoms that dissolve in the melt as a function of time. With  $N_c$  representing the total number of C atoms which have dissociated from the solid and have been transferred into the melt at time  $t$  and  $N_t$  the total number of MC steps, the dissolution rate can be computed throughout the duration of the dissolution process using the following equation:

$$\frac{dC}{dt} = \frac{N_c}{N_t} \quad (4)$$

The computed magnitudes of energy, dissolution rate, interfacial profile and widths depend on a large number of simulation variables, e.g., lattice size, contact area, interaction strengths etc. As simulation time is measured in units of Monte Carlo steps, no attempt was made to compare actual magnitudes of simulation results with the experimental values quoted in literature. Instead, the focus was on identifying various simulation trends.

### 3.1 Interfacial reactions

Fig. 2 shows a plot of the initial atomic profile of the graphite/Fe-C system across the basal plane. The initial carbon concentrations in the melt were chosen to be 0.0 wt% and 2 wt%. The layer  $Z = 80$  indicates the initial boundary between the solid and the melt. Fig. 2b shows the atomic distribution after simulation. Fe atoms appeared to have moved by up to 40 layers into graphite. On the other hand, C atoms moved right up to the melt boundary. The midpoint of C profile in graphite also moved back from  $Z = 80$  to  $Z = 55$ , indicating significant dissolution of graphite.

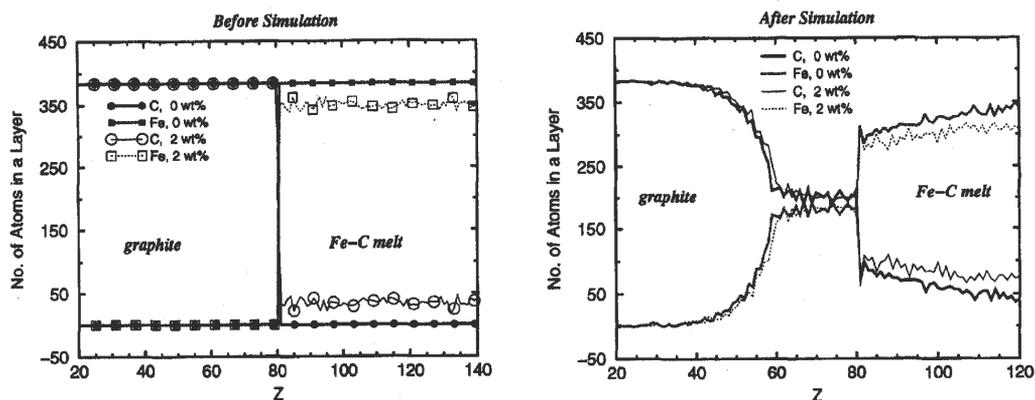


Fig. 2. Atomic distribution profile across graphite/Fe-C interface before and after simulation.  $Z$  corresponds to the layer number normal to the interface.

Once a C atom broke away from graphite, it was found to propagate easily in the melt. Fe atoms, however, found it difficult to penetrate solid graphite. The movement of a C atom in the melt involved the replacement of a repulsive Fe-Fe bond with a less repulsive Fe-C bond or an attractive C-C bond in place of a repulsive Fe-C bond. This process lowers the energy of the system and is therefore energetically favorable. On the other hand, the movement of an Fe atom replaces an attractive C-C bond with a repulsive Fe-C bond, a move not favored energetically. At  $T^* = 0$ , there should be no movement of Fe into graphite. On the other hand, a carbon atom once released from the graphitic substrate, can dissolve into the melt at all temperatures. Fig. 2b also shows a slight reduction in the width of the interfacial region with increasing carbon content of the melt.

The graphite dissolution rates across the basal plane have been plotted in Fig. 3a as a function of time for temperatures ranging from 1300°C to 1600°C. This plot can be divided in two regions. Region I, showing a sharp increase in dissolution rate, was restricted to a short period in the initial stages of contact between graphite and melt. C-C atom pairs across the basal plane have strong covalent bonds. It costs system energy to break these bonds and to replace them with repulsive Fe-C bonds. Once a few of these bonds get broken due to high temperature of the melt, it becomes relatively easier to release additional C atoms as some of these are now less strongly bound. During this short period, the dissolution rate of graphite is controlled by the slow reactions at the interface ( $k < K_m$ ).

The dissociation of C picked up after a short time. Increasing temperature also made the dissociation of C atoms easier and pushed region I to shorter times. In region II, the

dissolution was controlled by the mass transfer in the melt, which now becomes the slower process ( $K_m < k$ ). This represents the typical behavior observed in graphite dissolution experiments. The influence of temperature in region I was more significant as compared to region II. Due to generally higher activation energies, chemical reactions are more sensitive to temperature than mass transfer. This result supports the hypothesis that dissolution process is controlled by interfacial reactions in region I and by mass transfer in region II.

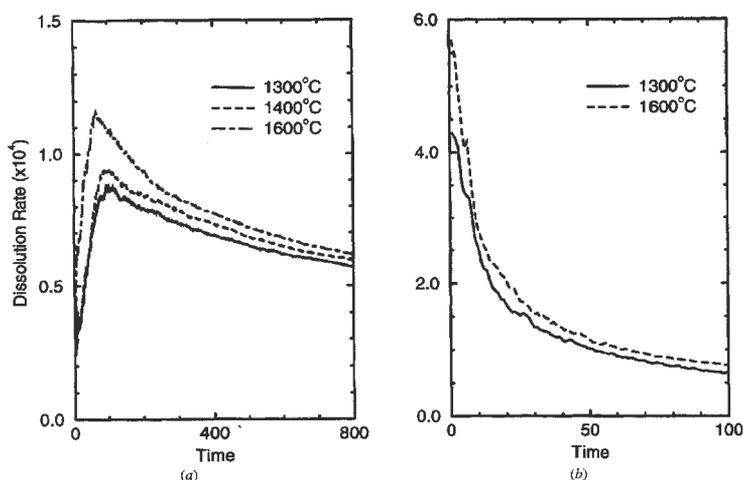


Fig. 3. Carbon dissolution rate as a function of time across (a) the basal plane and (b) the prismatic plane

The dissolution rates across prismatic planes are shown in Fig. 3b. The inter-layer bonding between C atoms is very weak and C atoms along this surface are very easily dissociated from graphite. Region I was not observed as there was no hindrance to C dissociation at the interface and the reactions at the interface were quite fast. The dissolution rates, which were much higher in the initial stages, were completely controlled by mass transfer in the melt.

A Monte Carlo simulation study was also carried out on the dissolution of a single graphite particle in Fe-C (0–4 wt%) melts in the temperature range 1400°C to 1600°C (Sahajwalla & Khanna, 2000). Using canonical ensemble, simulations were carried out as a function of particle size, carbon content of the melt and temperature. Simulation results showed that the dissolution of a graphitic particle does not take place layer by layer. The C atoms in the basal plane dissolved preferentially from the edges and iron liquid slowly moved in towards the centre. The atoms on a prismatic plane on the other hand dissolved all across the particle surface. Iron liquid penetrated deep in the particle and led to the formation of a broad interfacial region containing high concentrations of both C and Fe. The overall particle dissolution profile as a function of particle size, temperature and C concentration in the melt showed a good agreement with theoretical and experimental results. Even as carbon dissolution neared completion on a given surface, small islands of graphite could still be seen in a sea of melt.

### 3.2 Influence of sulphur

Initially sulphur atoms were distributed randomly and uniformly in the Fe- 4 Wt% S melt. Atomic distribution profiles for Fe, C and S atoms are plotted in Fig. 4 as number of atoms

in a given layer before and after simulation.  $Z=50$  represents the initial graphite/melt contact surface. When this melt was brought in contact with a block of graphite, various atoms started diffusing across the interface. While C atoms dissociated from the graphite block and dissolved in the iron melt, some Fe atoms also penetrated the graphite block (Sahajwalla & Khanna, 1999).

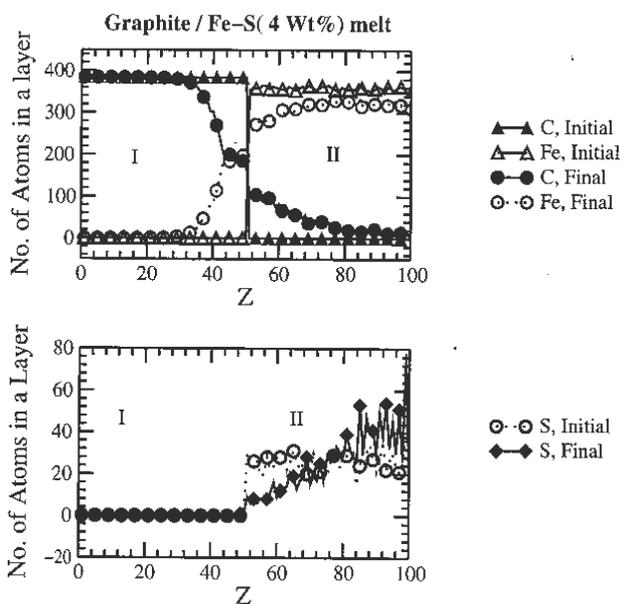


Fig. 4. Atomic distribution profile across graphite/Fe-S interface before and after simulation.  $Z$  corresponds to the layer number normal to the interface.

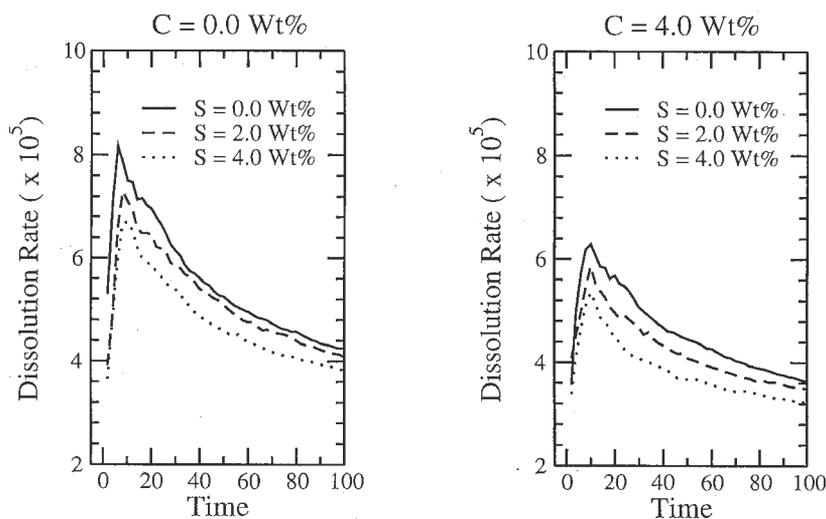


Fig. 5. Carbon dissolution rate as a function of time, across basal plane for a range of carbon and sulphur concentrations in the melt.

However there was no indication of S atoms penetrating the graphite block or blocking up the interfacial region. Instead these tended to move away from the interface and penetrate deep in the iron melt. This result is in excellent agreement with the EDS results of Wu et al (Wu et al., 2000) as they also did not find any evidence of a sulphur peak in the interfacial region in the initial stages of contact. There appeared to be well-defined regions of high C & low S close to the interface and low C & high S deep in the melt.

In Fig. 5, the dissolution rate of graphite is plotted as a function of time for a range of C and S concentrations of the melt. These results are for the basal plane of graphite in contact with hot melt. Due to strong covalent bonds between C atoms in the basal plane, dissociation of C atoms was quite slow and interfacial effects were rate controlling in the initial stages of contact. Dissociation rate picked up after some time and the carbon dissolution rate controlling mechanism changes over to mass transfer in the melt. However the presence of sulphur in the melt appeared to affect both these mechanisms.

The presence of sulphur in the melt affected the overall carbon dissolution rate adversely. This reduction takes place in the initial stages of contact and also when carbon concentration in the melt is close to saturation. In the initial stages of contact when interfacial effects were rate controlling, the presence of S led to a lowered carbon dissociation rate. This decrease was attributed to bond energy considerations and was not due to any interface blockage by sulphur atoms.

#### **4. Depletion of carbon from $\text{Al}_2\text{O}_3$ -C refractories into molten iron**

The atomic model of the graphite-alumina/liquid iron system was developed based on the atomic model of the graphite/liquid Fe-C system. These simulations were focussed on modelling the synthetic graphite-alumina/liquid iron system consistent with experimental observations in the system (Zhao & Sahajwalla, 2003). A validated atomistic model with optimum interaction parameters could then be used to systematically investigate the effect of various operating conditions such as temperature, melt turbulence, composition, on carbon dissolution and to provide guidelines for developing optimum refractories for steelmaking applications.

A key feature of this model was to include alumina molecules in the solid graphitic lattice. A molecule of alumina was represented as an inert, rigid unified group of five atoms, masking the fine detail regarding aluminum and oxygen constituents. This approach is quite common in computer simulation studies on complex molecules (Allen & Tildesley, 1987). As the molar volume for alumina ( $25.575 \text{ cm}^3$ ) is nearly five times the molar volume for graphite ( $5.298 \text{ cm}^3$ ) (Weast & Estle, 1982), a molecule of alumina was allocated five neighboring lattice sites as against a single site for a carbon and iron atoms. With iron being in liquid state in this study, the atomic size differences between C and Fe atoms were neglected.

Two identical blocks of dimensions  $L \times L \times L_d$ , with hexagonal sites, were placed in contact with each other (Fig. 6). Sites on the first block, representing alumina-graphite refractory were occupied by alumina and carbon distributed randomly. The relative concentrations of alumina and carbon atoms were governed by the refractory composition under consideration. All sites on the second block representing liquid steel were occupied by Fe atoms. In these studies, the basal plane of graphitic structure was placed in contact with the melt. This plane was preferred over the prismatic plane due to more significant interfacial reactions expected from this surface due to the relatively strong strengths of C-C bonds in the basal plane.

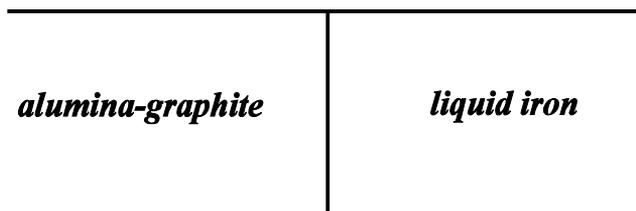


Fig. 6. A schematic representation of the simulation arrangement

The alumina–carbon interactions were assumed to be attractive in nature and the strength of the alumina–carbon interaction was varied over a range in an attempt to simulate the effect of binders commonly present in refractories. The alumina–iron interaction was assumed to be repulsive in nature due to their generally non-wetting behavior. While signs of these interaction parameters were estimated from fundamental considerations, their magnitudes were optimized and determined during these simulations. The interactions were chosen to be isotropic throughout the simulation lattice except for the C-C interactions, which were anisotropic due to weak van der Waals interactions along c-axis.

An attempt was made to incorporate the non-wetting between alumina and graphite through strong but finite repulsion between alumina and iron (Fig. 7a) and also by modifying the nature of interactions on the solid/liquid interface (Fig. 7b). Both approaches proved to be unsuccessful. A third attempt was then made to simulate non-wetting behaviour in terms of mutual exclusion of alumina and iron from their immediate neighbourhood (Fig. 7c). This approach coupled with the immobility of alumina showed a good fit to experimental results on the graphite–alumina/liquid iron system and helped in developing an atomistic model of the system.

In addition, the alumina interaction parameters were found to be redundant, having no effect on simulation results. The optimum interaction parameters for the graphite–alumina/liquid iron system were determined as:  $J_1(C-C) = -J$ ,  $J_1(C-Fe) = 0.6J$ ,  $J_1(Fe-Fe) = J$  and no interaction parameters for alumina. Turbulence in the melt was generated through additional atomic motion on melt sites. Results on the effect of melt turbulence on carbon dissolution from three refractory compositions are shown in Fig. 8a. With increased turbulence, there was an increase in the rate of carbon dissolution in the initial stages of contact for high carbon refractory mixtures (90%C and 70%C). However there was no significant change in their steady-state carbon content. This result can be understood in terms of increased rates of mass transfer in the melt caused by the melt turbulence leading to an improvement in the overall carbon dissolution rates. Melt turbulence did not have any effect on carbon dissolution from the refractory mixture containing 50%C.

The effect of temperature on carbon dissolution is shown in Fig. 8b. An increase in temperature from 1400°C to 1600°C would result in an increase in the transition probability  $W$  (Eq. 2) for Monte Carlo steps. While there was no noticeable change observed in carbon dissolution from refractory mixture containing 50%C, a substantial increase in carbon dissolution was observed for the mixture containing 90%C. A marginal improvement in carbon dissolution rates was observed for 70%C. While both temperature and melt turbulence had a significant influence on high carbon systems, these results clearly indicate that for refractory mixtures containing alumina in excess of 50%, carbon depletion from the refractory was not affected by the increased levels of melt turbulence or higher temperatures.

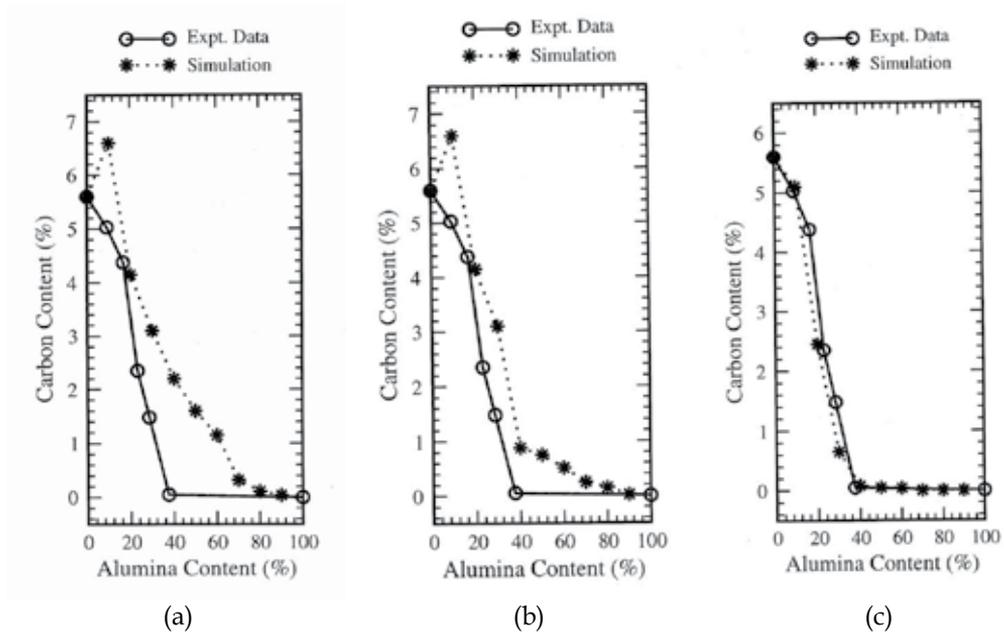


Fig. 7. Atomistic simulation results on graphite-alumina/liquid steel system. The non-wetting between alumina and liquid iron modelled as (a) strong but finite repulsion, (b) a modification of the nature of interactions on the solid/liquid interface, (c) mutual exclusion of alumina and iron from their immediate neighbourhood.

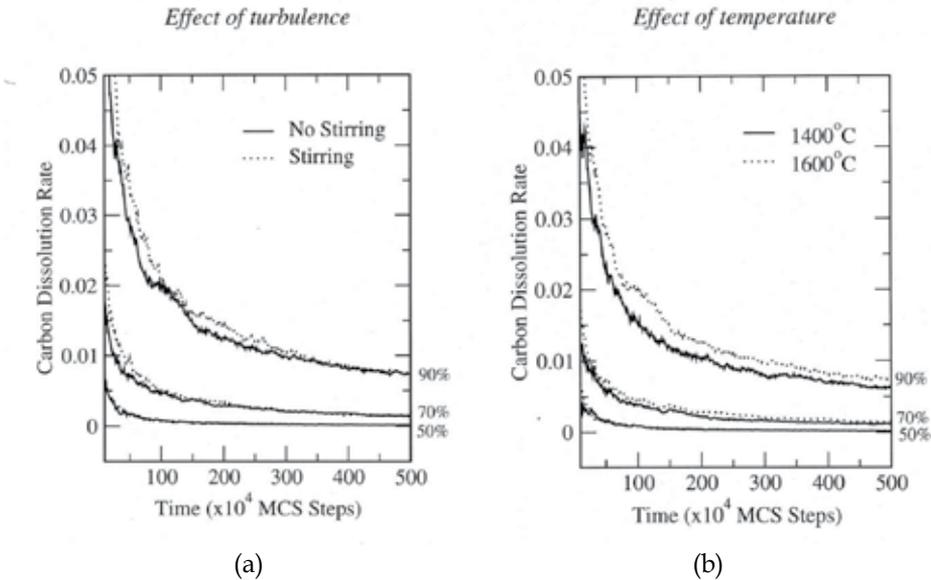


Fig. 8. Influence of operating parameters (a) the melt turbulence, and (b) temperature on carbon dissolution from three refractory mixtures. Concentrations on the right indicate the initial carbon concentration in the synthetic graphite-alumina refractory mixture.

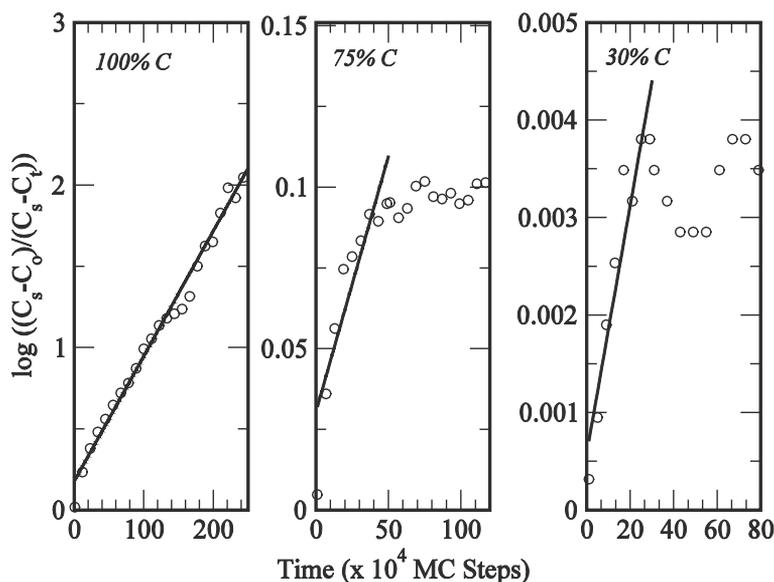


Fig. 9. Simulation results on  $\log((C_s - C_t)/(C_s - C_0))$  plots as a function of time for alumina-carbon refractories in contact with liquid iron at 1600°C.

In Fig. 9, we have plotted  $\log((C_s - C_t)/(C_s - C_0))$  as a function of time for three substrates (Sahajwalla et al., 2006). While a linear trend was observed for synthetic graphite (100% C), simulations from refractories containing 75% C and 25% C respectively did not show a linear correlation between  $\log((C_s - C_t)/(C_s - C_0))$  and time. Simulation results indicate that carbon dissolution in refractory mixtures was not governed by a first order kinetic process. Atomistic simulations and experimental results on these three systems have shown similar trends.

These results indicate that mass transfer in the melt was not a dominant rate controlling mechanism for alumina-carbon refractories. Poor wettability of alumina with liquid iron and its significant influence on inhibiting the penetration of liquid iron in the refractory matrix, and consequently a very limited contact between carbon and liquid iron was found to be the dominant mechanism through which carbon dissolution from refractories gets strongly suppressed. This study has important implications for a fundamental understanding of refractory behaviour at high temperatures and for further developments in commercial alumina-carbon refractories which use up to 35% C in their formulation.

## 5. Decarburisation reactions in molten Fe-C-S system: the role of free surfaces

Carbon levels in steel are controlled carefully through decarburization and refining processes (Mclean, 2006; Sain and Belton, 1978). During decarburization of continuously carbon-saturated liquid iron by  $\text{CO}_2$  between 1280°C and 1600°C, the presence of small amounts of sulphur (0.01 to 1 wt%) was found to significantly slow down the decarburization rate (Roddiss, 1973). It was also found that during the decarburization of iron melts containing 1-3 wt% C and up to 3 wt% S using an  $\text{O}_2$ -Ar mixture, an oxide film was consistently observed on the metal droplet surface. Increasing sulphur contents generally increased the tendency of surface oxide formation under otherwise similar conditions.

Gao et al (2000) carried out an in-depth investigation on the influence of sulphur content on the carbon boil phenomena and CO generation in Fe-C-S droplets (C= 4.2 wt%, S= 0.005-0.4 wt%, 1370-1450°C). Carbon boil is known to occur when the internal pressure of CO gas is high enough to exceed the surface energy of the metal drop. Times for carbon boil were found to range between 5 to 20 seconds; these were independent of sulphur up to a certain level and then decreased with a further increase in the sulphur content. During these decarburization studies, sulphur levels remained fairly unchanged throughout the duration of the experiment.

Existing theories were found to be inadequate in explaining key experimental results on the decarburization of steel, a very important aspect of steelmaking. Using ideal monolayer adsorption isotherms, Sain and Belton (1976) assumed that the surface activity effects of sulphur could be due to its occupying a majority of the surface sites thereby blocking the reaction. For high sulphur levels in the melt (~0.02 wt% S), these models predicted a complete coverage of the surface with sulphur and zero decarburization of the melt. A fresh look at the influence of sulphur on the kinetics of decarburization reactions was therefore required to develop a better understanding of reaction processes over a range of reaction conditions.

Computer simulations on thin films have reported that surface effects extend up to the top three atomic layers, where the surface characteristics are significantly different from the bulk (Binder et al., 1995; Shpyrko et al., 2005). It is quite likely that the surface of molten steel is multi-layered instead of a monolayer and could include several neighbouring layers close to the liquid surface. If sulphur atoms are present on a number of surface layers instead of the top monolayer, there is a strong likelihood for elements other than sulphur to be present on the surface thereby pointing to additional reaction pathways.

Atomistic computer simulations were carried on molten Fe-C-S system using isotropic atomic interaction parameters; free surfaces were characterised by a missing layer of atoms. Simulations were carried out as a function of melt carbon and sulphur concentration, temperatures and surface/volume ratios of the simulation cell. Monte Carlo simulations based on lattice gas models do not allow continuous motion of atoms; with various atoms occupying rigid lattice sites, complex distance dependent potentials were replaced by appropriate interaction parameters. Simulations were based on the bulk interaction parameters; surface interactions parameters between Fe, C, and S atoms were assumed to be identical in magnitude to the corresponding bulk interaction parameters. The simulation cell geometry is schematically shown in Fig. 10.

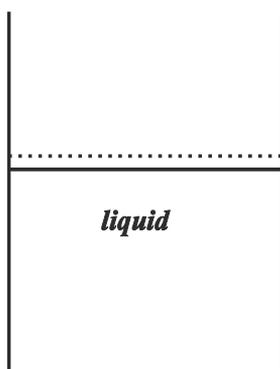


Fig. 10. Molten bath simulation cell configuration; dotted line represents the free liquid surface.

Turbulence in the liquid at high temperatures allowed a continuous/sporadic movement of atoms within the liquid. Periodic boundary conditions were not used for surface atoms. Another key difference between the surface atoms and those in the bulk was in the form of absent atomic planes/bonds, which could in turn result in major differences in overall interaction energies. The size  $L$  of the simulation cell was chosen to be quite large, ranging from 24 to 36; depth  $M$  was chosen to range between 10 and 300 to investigate bath configurations with different surface/volume ratios. The surface to volume ratio ( $S/V$ ) was computed as a dimensionless number by dividing the number of surface sites by the total number of sites in the simulation cell.

Computer simulations were carried out at 1400°C and 1500°C for melt carbon contents of 3.2 wt% and 4.2 wt%. Sulphur concentration in the melt typically ranged from 0 to 1.5 wt%. Simulation results for the configuration: 24x24x15 lattice (No. of sites in a layer =384),  $C = 4.2$  wt%, 1500°C, are shown in Fig. 11. Sulphur was found to predominantly concentrate in the top few layers with very little sulphur found in the interior bulk. The sulphur concentration was found to be highest in the second layer; this result was observed for all sulphur concentrations under investigation.

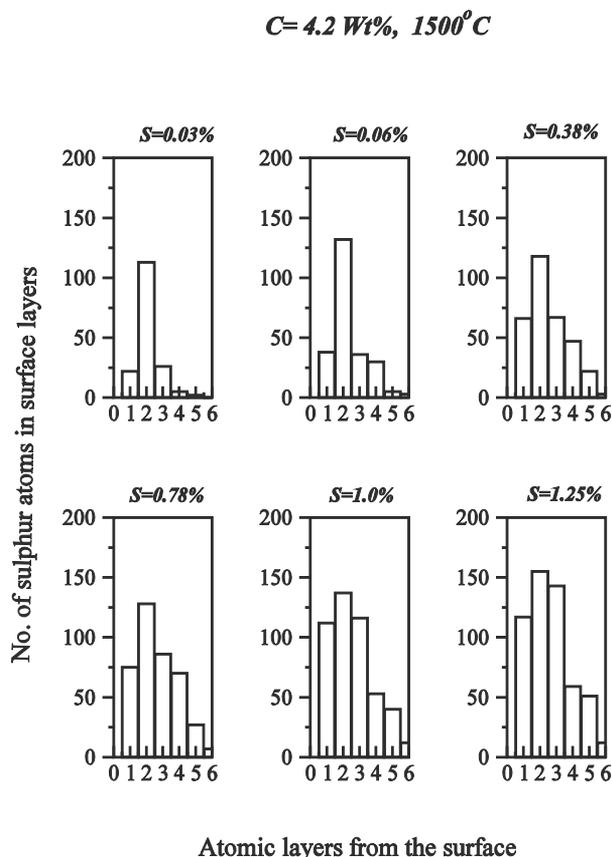


Fig. 11. Concentration of sulphur in the top five surface layers in a liquid bath configuration (24x24x15). Results for  $C = 4.2$  wt%,  $T = 1500^\circ\text{C}$  have been presented. Corresponding bulk sulphur values ranging from 0.03% to 1.25% have also been shown above individual bar plots.

With 1.25 wt% average sulphur concentration in the bulk melt, S atoms occupied 118 sites out of the 384 sites available on the top surface layer. This result clearly shows that a complete surface coverage with sulphur does not take place even when the number of available sulphur atoms far exceeds the number of surface sites available. Instead, these are distributed in a 3-5 atomic layer thick surface region leaving plenty of sites available for other atomic species. This trend, which appears to be a fundamental characteristic of the system, was observed for various carbon and sulphur levels, temperatures and bath depths. This finding is significantly different from the monolayer theory for the surface proposed by Sain and Belton (1978).

In Fig. 12 we report on the influence of surface/volume (S/V) ratio on the atomic concentration profiles of sulphur, iron and carbon in the top 5 surface layers. As significant variations in the melt carbon concentration did not have much influence on the sulphur distribution in the surface region, as a representative example we present specific results for an iron melt with C = 4.2 wt%, T = 1500°C. Total sulphur concentration in the melt was chosen to be 0.5 wt% (0.025 at. %). In the liquid bath configuration (24x24xM), S/V ratios ranging from 0.0667 to 0.2 were obtained by varying the bath depth M from 20 to 10. Most of the sulphur available in the melt was found to concentrate in the surface region; sulphur levels in the bulk liquid were found to be negligibly small.

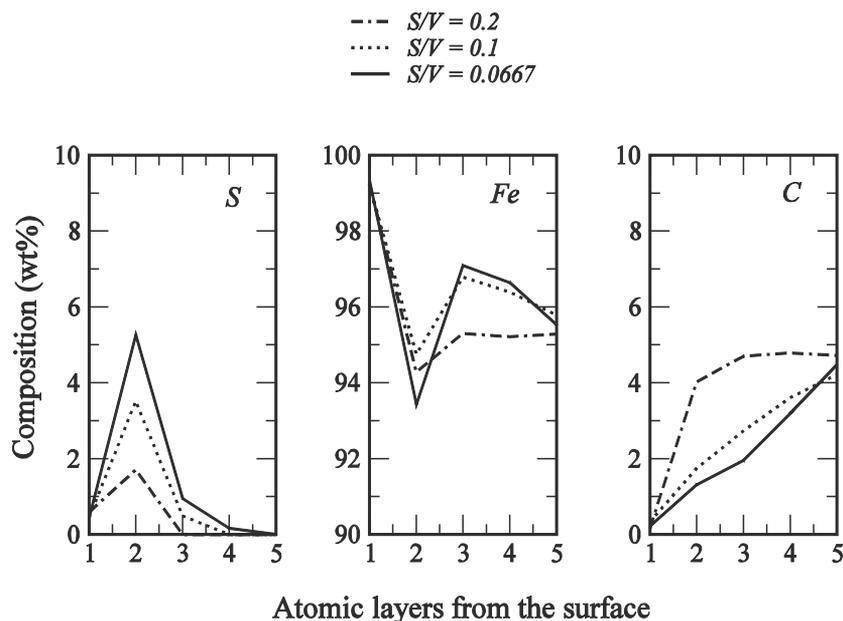


Fig. 12. The effect of surface to volume ratio on the atomic composition profiles in the surface region of liquid bath configuration (24x24xM). The depth M of the bath ranged from 10 to 20 layers for S/V ratios ranging from 0.2 to 0.0667.

In Fig. 13, we present the influence of S/V ratio in a few extreme scenarios. Very small S/V values were generated by significantly increasing bath depths. Bath depths ranging between 100 and 300 layers for a 24x24 bath corresponded to S/V ratios between 0.01 and 0.0033 respectively. Sulphur concentrations in the surface region were found to be a strong

function of S/V ratio. For the lowest S/V ratio investigated (0.0033), the sulphur concentration in the 2<sup>nd</sup> layer showed a sixty fold increase to 30 wt%. Carbon on the other hand was seen to be very low in the surface region. This result indicates that the influence of free surfaces on sulphur redistribution gets significantly reduced with increasing surface areas. For large exposed surfaces, the use of bulk concentrations might provide a reasonable basis for reaction rate computations. Surface active behaviour for sulphur becomes significant for relatively small exposed surfaces and then needs to be appropriately taken into account.

It is important to note that the surface to volume (S/V) ratio has been defined in these simulations as a dimensionless number. For establishing a correspondence between simulation and experimental results and to convert surface/volume ratio in units of  $m^{-1}$ , S/V ratios need to be scaled by a factor of  $3.10^9$  ( $=1/d$ , where 'd' is inter-planar spacing). The simulation results presented above represent systems with large surface to volume ratios, typically observed for micro (or nano)-scale systems, e.g., small liquid droplets. In a macro-scale system, e.g., (1x1x0.001 m) bath, the S/V ratio will be  $10^3 m^{-1}$ ; the number of surface sites will be a small fraction of the total number of atomic sites and the effect of free surfaces will be much reduced.

### Deep Molten Bath

***C= 4.2 Wt%, S= 0.5 Wt%, 1500° C***

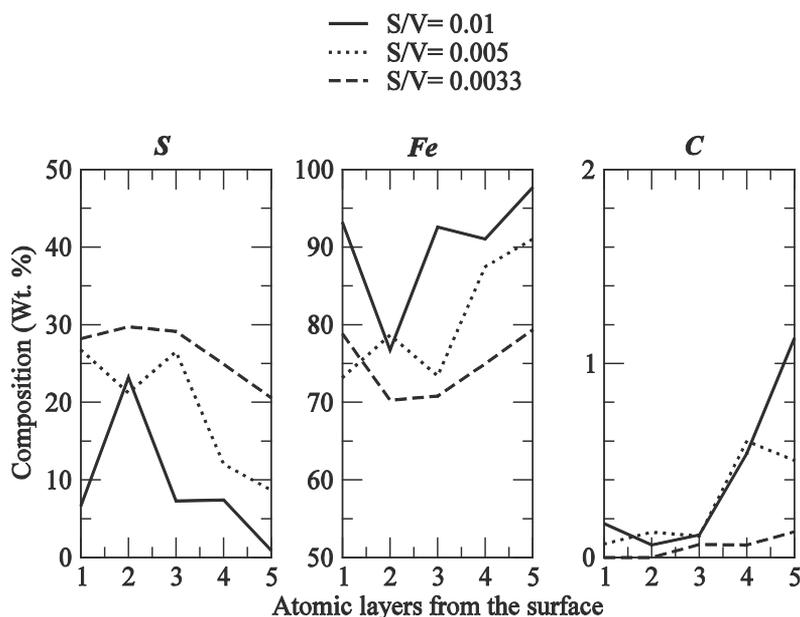


Fig. 13. The effect of surface to volume ratio on the atomic composition profiles in the surface region of liquid bath configuration (24x24xM); C=4.2 wt%, 1500°C. The depth M of the bath ranged from 100 to 300 layers for S/V ratios ranging from 0.01 to 0.0033.

The system was investigated further to probe the influence of temperature and carbon concentration. Temperature was varied from 1400°C to 1600°C, and the carbon

concentration in the melt from 1.2 wt% to 5.2 wt%. Total S was maintained at 0.5wt% (0.025 at. %), and the bath configuration had an S/V ratio of 0.0667. Sulphur was found only in the surface region with very small levels sulphur detected in the bulk region at all temperatures and carbon levels under investigation. Bulk sulphur levels were therefore not reported. In the surface region, there was not much influence of temperature or of carbon concentration on sulphur levels. No well-defined trends could be observed. Carbon levels in the surface region were however significantly lower than the corresponding bulk values. For a 5.2 wt% C bulk concentration, the surface region showed ~ 2% carbon and the bulk region showed a slightly higher value of 5.6 wt%. Temperature once again had only a marginal influence.

Iron levels in the bulk region were marginally higher than their corresponding levels in the surface region, which was caused by several of the surface sites being preferentially occupied by sulphur atoms. Iron concentration was higher than 94% across the entire simulation zone. This result indicates that the liquid surface was not completely depleted of iron; iron present in the surface region could participate in a number of surface-based reactions previously considered to be blocked by the sulphur monolayer. There was little influence of temperature. The continuous decrease in iron levels observed with increasing carbon was probably due to overall changes in the bulk metal composition; the presence of free surfaces did not have any role in this decrease.

The influence of sulphur on the re-distribution of carbon within the liquid is an important and novel finding. A high surface accumulation of sulphur automatically resulted in low surface concentrations of carbon. As sulphur was concentrated in the top few layers, the concentration of carbon was significantly reduced in the surface region, and was slightly higher in the bulk region. For 3 wt% C in the bulk, ~1 wt% C was observed in the surface region along with ~3.5 wt% C in the bulk region. Surface decarburization and carbon boil phenomena will depend strongly on the carbon concentration in the bulk and in the surface regions. In contrast to the results observed for sulphur, increasing surface areas led to much higher concentrations of carbon closer to the surface, which can be construed as a potential nanoscale effect for liquid steels. Atomic redistribution of carbon caused by the presence of free surfaces could therefore have a significant influence on the operational efficiency of direct ironmaking smelting technologies.

Another key result was finding up to 99% surface sites being occupied by Fe atoms. Fe-Fe interaction parameters are repulsive in nature with iron being in the molten state at these temperatures. The system tends to lower its energy by concentrating repulsive bonds on free surfaces; both C and S had moved away from the outermost layer. However since S is a surface active element and has a strong attractive interaction with Fe, S tended to concentrate in the second layer. Another key finding was on the influence of sulphur on the re-distribution of carbon within the liquid. Carbon and sulphur atoms tend to displace one other in molten steel; a high surface accumulation of sulphur will automatically result in low surface concentrations of carbon. As sulphur was concentrated in the top few layers, the concentration of carbon was significantly reduced in the surface region, and was correspondingly higher in the bulk region. The presence of free surfaces therefore resulted in a major redistribution of various atomic species.

Observed simulation findings are consistent with the formation of iron oxide layer on Fe-C-S droplets in oxidising atmospheres. Iron needs to be a major reacting species present on the surface for oxidising reactions to take place. Secondly, since a major proportion of sulphur

was present in the second layer and not in the top outermost layer, it would significantly reduce its direct interactions with oxidising gases. This would help minimise loss of sulphur to the environment and could account to a certain extent for the negligible loss of sulphur observed experimentally.

For decarburisation reactions under CO<sub>2</sub> atmospheres, the addition of small amounts of sulphur resulted in a significant reduction of decarburization rate. This experimental result was explained by Sain and Belton (1978) by assuming that S atoms blocked most of the surface sites and only 1.4% active sites were available for other reactions. Based on our simulation results, we suggest an alternative approach. Due to mutual displacement between C and S, the presence of S in the surface region would automatically move C away from the surface thereby significantly reducing its direct contact with CO<sub>2</sub> gas. The reduction in decarburization rate can also be explained by C atoms moving away from the surface; a complete blockage of surface by sulphur may not be necessary for slowing down reactions. The results shown in Fig. 11 to 13 have shown sulphur concentrating in the top few layers, with highest sulphur levels being observed in the second layer. Secondly, we have observed that the top layer always contained small amounts of carbon as well. This finding could account for the observed residual decarburisation without the need for arbitrarily blocking reactive sites through atomic misfit or orientation considerations.

Simulation results suggest a novel approach is required to enhance our current understanding on the influence of surface-active elements. Theoretical models using the concept of reactive sites suggest that surface activity effects of sulphur may be due to its occupying a majority of surface sites; a complete coverage of surface with sulphur has been suggested for sulphur levels ~ 0.07 wt%. Although our results also indicate sulphur concentrating close to the surface, these atoms are present in several top layers instead of a monolayer. Simulations indicate that a complete surface blockage by sulphur is not likely and opens up the possibility for surface reactions with other atomic species such as Fe.

## 6. Concluding remarks and future research directions

### 6.1 Key conclusions

1. Using a rigid graphitic hexagonal lattice, an atomistic model has been developed for the graphite/iron melt interfacial region. This model was validated using the saturation solubility of C in Fe melts and by the retarding effect of sulphur on C solubility. A Monte Carlo algorithm was developed using canonical ensemble for investigating the dissolution behaviour of graphite and the reactions taking place at the graphite/melt interface. A contact between graphite and melt led to the dissociation of C atoms from the graphite lattice and their subsequent diffusion in the melt. Some Fe atoms also penetrated the graphite lattice.
2. The rate of mass transfer of C atoms in the melt was found to be slower than the corresponding dissociation rate at the interface. This led to a build up of C atoms at the interface. A higher initial C concentration in the melt resulted in a smaller width of the interfacial region. The dissolution of graphite did not take place layer by layer. Instead, a broad interfacial region containing high concentrations of C and Fe was formed in the top layers of graphite; the initial shape of the dissolving particle remained fairly unchanged over extended periods of contact.

3. Strong C-C bonds in the basal planes of graphite were found to hinder the dissociation of C atoms from this surface. During the initial stages of contact, interfacial reactions on this interface were very slow and controlled the dissolution process. The dissociation of C atoms became easier and faster with extended contact. Across prismatic planes, the bonding between C atoms was very weak and the atoms could be easily dissociated. Interfacial reactions did not play an important role across this interface.
4. An atomistic model was developed for the synthetic graphite-alumina/liquid iron system. An attempt was made to incorporate non-wetting between alumina and graphite through strong but finite repulsion between alumina and iron and also through modifying the nature of interactions on the solid/liquid interface. Both approaches proved to be unsuccessful. A third attempt was then made to simulate the non-wetting behaviour in terms of mutual exclusion of alumina and iron from their immediate neighbourhood. This approach coupled with the immobility of alumina was able to simulate a number of key experimental results on the graphite-alumina/liquid iron system.
5. Alumina interaction parameters were generally found to be redundant having no effect on simulation results. Further simulations were carried out to determine the influence of melt turbulence and temperature on carbon dissolution from synthetic graphite-alumina/liquid iron system. While both temperature and melt turbulence had a significant influence on high carbon systems, refractory mixtures containing alumina in excess of 50% were not affected by increased levels of melt turbulence or higher temperatures. Poor wetting in the interfacial region appears to be the most important factor affecting carbon depletion from refractories.
6. Simulation results have clearly shown that the free surface of molten Fe-C-S can not be approximated to a mono atomic layer. Instead, a surface may be represented by a 3-5 atomic layer thick surface region, where the atomic concentration profiles are significantly different from the bulk liquid. The outermost layer of molten Fe-C-S system was found to be predominantly occupied by Fe atoms; a result that would explain the formation of an iron oxide layer on the liquid surface in oxidising atmospheres. Highest concentration of sulphur atoms in close vicinity to iron atoms in the second layer could account to some extent for the negligible loss of sulphur during the reactions. These experimental results were not previously well explained by the existing theories.
7. The presence of sulphur had a significant influence on the redistribution of carbon in the surface and bulk regions. The carbon concentration in the bulk region was slightly higher than the bulk carbon levels; the surface region on the other hand showed much lower levels of carbon. Increasing surface areas led to much higher concentrations of carbon in the surface region. This result is of great significance for surface decarburization reactions and carbon boil phenomena in direct smelting technologies; the impact of free surface therefore needs to be taken into account for optimizing process efficiencies.
8. The surface-active sulphur was found to concentrate preferentially in the surface region; very little sulphur was found in the bulk region of liquid. With 1.5 wt% S in the bulk melt, S atoms occupied ~30% sites available on the top surface layer. This result clearly shows that a complete surface coverage with sulphur does not take place even when the number of available sulphur atoms exceeds the number of surface sites

available. Therefore the ideal monolayer theory, wherein surface-active elements block all the reactive sites on the surface, does not provide an accurate representation of liquid surface in the presence of surface active species.

## 6.2 Future research directions

The characterization of a liquid surface by the missing atomic layers is only a partial representation of surface effects in liquids. Due to surface tension effects, bonds are likely to be stretched along the surface and compressed in a direction normal to the surface. Surface interaction parameters need to be modified and their deviations from bulk interactions determined. These modifications are rather complex and nontrivial in nature, and need to take into account the thickness of the surface layer; the nature, strength and gradients of individual pair potentials. Simulations are currently underway to determine optimum surface interaction parameters and associated surface thermodynamics.

Due to the importance of iron in a number of fields such as magnetism, earth sciences and iron/steel making, a very large number of 'r' dependent potentials have been reported in the literature that mimic both cohesive and repulsive energetics of materials. Although it is well understood that such models are empirical in nature and fitted to specific material characteristics and applications, these are widely used in atomistic investigations due to their computational efficiency and practical applicability. One of the most widely used schemes is the embedded atom method (EAM) which has been successfully implemented to describe the cohesive properties of metals and the de-cohesive role of impurities such as hydrogen (Daws & Baskes, 1983). The effects of boundary conditions on the melting process of an ensemble of bcc iron atoms will be investigated by performing the Monte Carlo simulation and using a Lennard-Jones and other distance-dependent pair potentials. The temperature-energy phase diagrams will be used to establish correspondence with iron allotropic transitions. The stability of melting process under fixed wall and free surface condition will be energetically/spatially analysed along with effects of short range order.

We also aim to determine key aspects of surface thermodynamics, bonding network, the influence of surface-active elements, pathways and mechanisms for surface reactions through atomic level investigations on the surface/sub-surface region of molten steel (1550-1650°C). Using a novel approach, key strategic ingredients in the surface reactions of Fe-C-S system, atomic concentration profiles and bonding networks will be determined for the first time with a view to developing a reliable atomistic characterization of the liquid steel surface region. The surface thermodynamical behaviour of molten Fe-C-S system will be determined through computations of surface interaction parameters, free energy and surface tension using an iterative procedure. High temperature experimental measurements will be used to supplement, provide feedback and to validate simulation results. Atomic concentration profiles and reaction pathways will then be determined over a wide concentration/and temperature range followed by a critical analysis and implications for steelmaking reactions.

## 7. References

Binder, K.; Lebowitz, J.L.; Phani, M.K. & Kalos, M.H. (1981). Monte Carlo study of the phase boundary of binary alloys with face centred cubic structure. *Acta Metall.*, Vol 29, pp. 1655-65.

- Binder, K.; Landau, D.P. & Ferrenberg, A.M. (1995). Thin Ising films with competing walls: A Monte Carlo study. *Phys. Rev. E*, Vol. 51, No. 4. Pp. 2823-2838.
- Chen, C.; Lin, C & Chen, S. (2000) Kinetics of synthesis of silican carbide by carbothermal reduction of silicon dioxide. *British Ceramic Transactions*, Vol 99, No. 2, pp. 57-62.
- Chipman, J. (1972), Thermodynamics and the phase diagram for the Fe-C system. *Metall. Trans.*, 1972, Vol. 3, pp. 55-64.
- Coller, M.L.F; Sahajwalla, V.; Taylor, I.F. & Wright, J.K. (1994). Simulation of carbon dissolution from non-graphitic sources into iron melts. *6<sup>th</sup> AUSIMM extractive metallurgy conference*, Brisbane. Book Series: Australian Institute of mining and metallurgical publications. Vol. 49, No. 4. pp. 287-294.
- Cusack, B.L.; Hardie, G.J. & Burke, P.D. (1991) HIs melt - 2nd Generation direct smelting, *Second European Ironmaking Conference*, Glasgow.
- Daw, M.S. & Baskes, M.I. (1983) Semiempirical, quantum mechanical calculation of hydrogen embrittlement in metals. *Phys. Rev. Lett.* Vol. 50, pp. 1285-1288.
- Ericsson, S.O. & Melberg, P.O. (1981). Influence of sulphur on the rate of carbon dissolution in liquid iron. *Scandinavian Journal of Metallurgy*. Vol. 10. pp. 15-18.
- Gao, K.; Sahajwalla, V.; Sun, H.; Wheatley, C. & Dry, R. (2000) Influence of sulfur content and temperature on the carbon boil and CO generation in Fe-C-S drops, *ISIJ International*, Vol. 40, No. 4, pp. 301-08.
- Grigoryan, V.A. & Karshin, V.P. (1972). Effect of surface-active additives on the kinetics of dissolution of graphite in liquid iron. *Izvest Akad Nauk SssR Metally*. Vol. 1, pp. 209-214.
- Guillemet, A.F.; Hillert, M.; Jansson, B. & Sundman, B. (1981). An assessment of the Fe-S system using a two sublattice model for the liquid phase. *Metallurgical and Materials Transactions B*, Vol.12, pp. 745-754.
- Jordan, A.S. (1979). *Calculation of Phase Diagrams and Thermochemistry of Alloy Phases* (edited by Chang, Y.A. & Smith, J.F.), The Metallurgical Society of AMIE, New York, 1979, pp. 100-102.
- Keogh, J.V.; Hardie, G.J.; Philip, D. K. & Burke, P.D. (1991). HIs melt Process Advances to 100,000 t/y plant. *Ironmaking Conference Proceedings, ISS-AIME 50*, pp. 635-649, Washington DC.
- Khanna, R. & Sahajwalla, V. (1999) Monte Carlo Simulation of Phenomena at Solid Graphite/Fe-C melt Interface, *Physica Status Solidi B*, Vol. 213, No. 1, pp. 47-58.
- Kitchener, J.A.; Bockris, O.M. & Liberman, D. (1948) The activity of sulphur in liquid iron: the influence of carbon. *Faraday Society Discussions*, Vol. 4, pp. 49-61.
- Kosaka, M. & Minowa, S. (1968). On the rate of dissolution of carbon into molten Fe-C alloy. *Transactions ISIJ*. Vol. 8, pp. 392-400.
- Lacaze, J. & Sundman, B. (1991). An assessment of the Fe-C-Si system. *Metall and Mater Trans. A*, 1991, Vol. 22, pp. 2211-17.
- McLean, A. (2006) The science and technology of steelmaking- measurements, models and manufacturing. *Metallurgical and Materials Transactions B*, Vol. 37B, pp. 319-332.
- Mori, T., Fujimura, K. & Kanoshima, H., (1963) Effects of Aluminium, Sulphur, and Vanadium on the Solubility of Graphite in Liquid Iron, *Mem. Fac. Eng. Kyoto Univ.*, Vol. 25, pp. 83-105.

- Morris J. & Williams, A. (1949). The effect of silicon on the activity of sulphur in liquid iron. *Trans. Amer. Soc. Metals*. Vol. 41, pp. 1425-1440.
- Moriss, J. & Buehl, D. (1950) Activity of sulphur in Fe-C alloys. *Transactions AIME*, Vol. 188, pp. 317-322.
- Mourao, M.B.; Murthy, G.G.K. & Elliot, J.F. (1993). Experimental investigations of dissolution rates of carbonaceous materials in liquid iron-carbon melts. *Metallurgical and Materials Trans.* Vol. 24 B, pp. 629-638.
- Ohtani, H. & Nishizawa, T. (1986). Calculation of Fe-C-S ternary phase diagram. *Trans. ISIJ*, Vol. 26, Vol. 9, pp. 655-663.
- Olivares, R. (1996). The effect of sulphur on the dissolution of graphites and carbon in liquid iron-carbon alloys. PhD Thesis, University of New Castle, Australia.
- Orston, S. & Oeters, F. (1988). Behaviour of Coal particles blown into liquid iron. *W.O. Philbrook Memorial Symposium Proceedings*, Iron and Steel Society, Ontario, pp. 27-38.
- Roddis, P. G. (1973) Mechanism of decarburisation of iron-carbon alloy drops falling through an oxidising gas. *J. Iron Steel Inst.*, Vol. 211, No. 1, pp.
- Sahajwalla, V.; Taylor, I.F.; Wright, J.K. & Hardie, G.J. (1994) Dissolution of carbon into iron melts- the new direct ironmaking perspective. *Metallurgical processes for the early twenty-first century*. Vol 1- Basic principles, pp. 715-730.
- Sahajwalla, V. & Khanna R. (1999) Dissolution Behaviour of Particulate Graphite in Fe-C melts: A Monte Carlo Simulation Study, *Scand. J of Metallurgy*, vol. 29, pp. 114-120.
- Sahajwalla, V. & Khanna, R. (1999). Influence of sulphur on the solubility of graphite in iron melts: a Monte Carlo simulation study, *Acta Materialia*, Vol. 47, No. 3, pp. 793-800.
- Sahajwalla, V. & Khanna R. (2000) A Monte Carlo simulation Study of Dissolution of Graphite in Iron-Carbon melts, *Metallurgical and Materials Trans.* Vol. 31 B, pp 1517-1525.
- Sahajwalla, V. & Khanna R. (2002) Influence of Sulphur on the Solubility of Graphite in Fe-C-S melts: Optimization of Interaction Parameters, *Acta Materiala*, Vol. 50, pp 663-671.
- Sahajwalla, V.; Khanna, R.; Kapilashrami, E. & Seetharaman, S. (2007) Depletion of carbon from Al<sub>2</sub>O<sub>3</sub>-C mixtures into liquid iron: Rate Controlling Mechanisms, *Canadian Metallurgical Quarterly*, Vol. 46, No. 1, pp. 25-32.
- Sain, D.R. & Belton, G.R. (1976). Interfacial reaction kinetics in the decarburisation of liquid iron by carbon dioxide. *Metallurgical and Materials Transactions*. Vol. 7, pp. 235-244.
- Sain, D.R. & Belton, G.R. (1978). The influence of sulphur on Interfacial reaction kinetics in the decarburisation of liquid iron by carbon dioxide. *Metallurgical and Materials Transactions*. Vol. 9, pp. 403-407.
- Sasai, K. & Mizukami, Y. (1995). Improvement of life inversion nozzles with gas injection. *ISIJ International*. Vol. 35. No. 9, pp. 1072-1078.
- Schei, A.; Tuset, J.K. & Tverit H. (1988), *Production of High Silicon Alloys*, Tapir Forlaug, Trondheim.
- Shpyrko, O.G.; Grigoriev, A.Y.; Streitl, R.; Pontoni, D.; Pershan, P.S.; Deutsch, M.; Ocko, L. & Meron M, Lin B. Atomic-Scale Surface Demixing in a Eutectic Liquid BiSn Alloy. *Phys. Rev. Lett.* 2005; 95: 106103-106107.

- Van Vlack, L.H. (1989). *Elements of Materials Science and Engineering*, 6<sup>th</sup> ed., Edison-Wesley Publishing Co., New York, pp. 133-34.
- Wright, J.K. & Taylor, I.F. (1993) Multiparticle Dissolution Kinetics of Carbon in Iron-Carbon-Sulphur Melts. *ISIJ International*. Vol. 33, No. 5, pp. 529-538, ISSN: 1347-5460.
- Wu, C.; Wiblen, R. & Sahajwalla, V. (2000) Influence of ash on mass transfer and interfacial reaction between natural graphite and liquid iron, *Metallurgical and Materials Transactions B*, Vol. 31B, pp. 1099–1104.
- Zhao, L. & Sahajwalla, V. (2003) Interfacial phenomena during wetting of graphite/alumina mixtures by liquid iron, *ISIJ International*, Vol. 43, No. 1, pp. 1-6.

# GCMC Simulations of Gas Adsorption in Carbon Pore Structures

Maria Konstantakou, Anastasios Gotzias,  
Michael Kainourgiakis, Athanasios K. Stubos and Theodore A. Steriotis  
*National Center for Scientific Research Demokritos*  
15310 Ag. Paraskevi, Athens,  
Greece

## 1. Introduction

The development of algorithms about 50 years ago gave to the scientific community an extremely powerful tool, allowing the utilization of digital computers to simulate and predict the thermodynamic, structural and dynamic properties of bulk fluids. The most widely used simulation methods for molecular systems are Monte Carlo and Molecular Dynamics. These methods provide a link between microscopic (molecular level) and macroscopic behavior, by simply evaluating numerically fundamental equations of statistical mechanics. Their great advantage is the ability to treat large systems having big number of molecules in relatively small time.

In the beginning, due to the limited computer power and availability, the studies were focused on the investigation of properties of the simplest possible fluid model (hard sphere) in the bulk state in two and three dimensional systems. The very first computer simulation of a liquid was carried out at the Los Alamos National Laboratory by Metropolis in 1953 (Metropolis et al., 1953). The technique employed, i.e. "Metropolis Monte Carlo", uses a Markov chain to generate a series of molecular configurations (microstates) by stepping from one configuration to the next with appropriate probability rules (Allen & Tildesley, 1987; Frenkel & Smit, 1996). Finally, the requested quantities (thermodynamic properties of interest) can be averaged over the configurations.

The rapid increase of computer resources promoted the scientific interest to explore the behavior of more complex systems, such as the thermodynamic properties of gas molecules confined in small cavities. In this case, beyond the calculation of the interactions between a molecule and the surrounding fluid, the solid adsorbent must be accurately reproduced and adsorbate-adsorbent interactions must also be considered. The first theoretical studies aiming to investigate the phenomenon of gas adsorption in porous solids were published in the 1960's (Alder & Wainwright, 1960). Since then, a large number of theoretical studies has been published, comparing experimental and simulation results and thus evaluating the accuracy of the models (gas-solid and gas-gas interactions) employed. Nowadays, computer simulation of gas adsorption is considered to be an extremely useful tool for explaining and analyzing experimental results but also for predicting gas solid equilibria.

Adsorption in porous materials is utilized in several industrial (food, pharmaceutical and petrochemical industry) and geophysical applications, for pollution control and

environmental protection, mixture separation, water purification and gas storage (Ruthven, 1984). Moreover, gas adsorption measurements are widely used in materials science as a reliable method for the characterization of porous materials (Gregg & Sing, 1982; Lowell & Shields, 1991). The better understanding of the detailed mechanism that takes place during these operations is essential for designing improved processes. The materials that are commonly used in these fields are crystalline, ordered, and amorphous porous materials, such as zeolites, amorphous silica and alumina, activated carbons, metal-organic frameworks (MOFs) and other advanced materials. The large surface areas of these materials and the confinement offered by their extended pore network enhance their catalytic, sorptive and separation activity.

Porous materials are classified by the International Union of Pure and Applied Chemistry (IUPAC) in micropores with pore diameters less than 2 nm, mesopores having pore widths between 2 and 50 nm and macropores with pore diameters greater than 50 nm (Everett, 1972). The pore width is defined as the diameter ( $D$ ) in the case of cylindrical pores or as the distance between opposite walls ( $H$ ) in the case of slit-shaped pores. Modern industrial and technological needs has led materials science (and vice versa) towards the development of novel materials having extremely small pore widths. The size of these pores is approaching few molecular diameters and materials with such pore systems reveal a wide range of properties, that differ significantly from mesopores and big micropores. Adsorption depends strongly on the structural properties of the adsorbent material, e.g. the specific surface area, the porosity and the pore dimensions. In general, the existence of a large specific surface area and of an extensive number of readily accessible small sized pores is desirable as in pores of molecular dimensions the adsorbent field is further intensified due to the overlapping solid wall potentials, resulting in enhanced adsorption capacity.

The pore filling mechanism is sufficiently described for the case of mesopores and macropores. The Kelvin equation is a purely thermodynamic model that applies at subcritical temperatures and relates the relative pressure ( $P/P_0$ ) at which capillary condensation occurs to the pore width (Gregg & Sing, 1982). However, the equation fails to apply in micropores, mainly because, a "real" adsorbate phase of molecular dimensions cannot be defined. Molecular models, as for instance the Density Functional Theory (DFT) (Seaton et al., 1989; Lastoskie et al., 1993; Aukett et al., 1992; Ravikovitch et al., 1995; Sosin & Quinn 1995; Scaife et al., 2000; Jagiello & Thommes 2004; Nguyen & Bhatia 2004) and the Monte Carlo (MC) technique (Nilson et al., 2003; Do & Do 2005; Nguyen et al., 2005), can offer a more comprehensive representation of the pore filling process. DFT is computationally less demanding and can provide an accurate description when dealing with simple fluids (spherical molecules) and simple geometries. However, as the precision of the microscopic models depends principally on the truthful representation of the molecules, including partial charges and atom sites, MC has been established as an efficient alternative approach. The method can give valuable information concerning the densification process in nanopores and moreover details about the packing structure of the gas molecules inside the pore can be extracted from the local density and molecule orientation profiles (Samios et al., 1997; Samios et al., 2000).

The MC method is widely applicable in carbon materials with large number of studies referring to the adsorption of different gases like hydrogen, carbon dioxide, methane, nitrogen, argon etc., for several applications. For example, during the last decade, the scientific community exhibited considerable interest about physical adsorption of hydrogen on various carbon based nanoporous materials. Likewise, adsorption in porous solids might

provide a useful tool for handling the challenging environmental issue of the high CO<sub>2</sub> atmospheric concentrations. The activities undertaken involve CO<sub>2</sub> separation, collection and finally storage in geologic formations (such as oil and gas fields, coal beds and saline formations). Thus, the knowledge of the behavior of CO<sub>2</sub> molecules during sorption in confined spaces like nanopores is of paramount importance. On the other hand, gas adsorption is used widely for the characterization of porous carbons, in terms of pore size distribution (PSD). The method combines experimental and simulated adsorption isotherms of gases like N<sub>2</sub>, CO<sub>2</sub>, Ar and lately H<sub>2</sub> (or combinations of them), in order to calculate the optimal distribution of the pore sizes of a material.

In this work the Grand Canonical Monte Carlo (GCMC) method is employed for the study of the adsorption behavior of H<sub>2</sub>, CO<sub>2</sub> and N<sub>2</sub> in three different carbon structures (slits, tubes and cones). Initially, the construction of solids, the representation of the gas molecules and the modeling of all the types of interactions are described. The significance of choosing the right potentials is emphasized, by giving an example about the H<sub>2</sub> quantum contribution at low temperatures. The influence of the pore size and geometry, as well as the temperature dependence is examined while the GCMC method is also used to study the packing structure of the gas molecules in the pores (local density and orientation profiles) as the temperature or pressure change. Finally, the simulated results are employed for the determination of pore size distribution of carbon porous materials.

## 2. Simulation model

Depending on the thermodynamic equilibrium properties sought a variety of Monte Carlo ensembles is available (Allen & Tildesley, 1987). The GCMC method describes a collection of microscopic systems of equal volumes in contact with a heat bath and a particle reservoir, i.e. the systems have fixed volume ( $V$ ), temperature ( $T$ ) and chemical potential ( $\mu$ ) (Nicholson & Parsonage, 1982). Each microscopic system (microstate) is literally an identical simulation box containing a reliable representation of the pore under investigation and a unique configuration of adsorbate particles determined by the applied  $\mu$  and  $T$ . As under these conditions GCMC permits fluctuations in density and energy the microstates are heuristically sampled and the averages of the fluctuating quantities are evaluated. The adsorption isotherm is then expressed as the output density (or the average number of adsorbate molecules) versus chemical potential ( $N=f(\mu)$ ) at a fixed temperature. The generation of different microstates is based on a Markov chain process, i.e. from any given molecular configuration a new one is generated by random insertion, deletion or displacement of an adsorbate molecule. If molecules are not spherical all moves are accompanied by random rotation (Marsaglia algorithm, Allen & Tildesley, 1987). Microstates are accepted with a probability that depends on the energy difference between the new (trial) and the old (current) configuration. According to the Metropolis sampling scheme random displacement is accepted with a probability  $p_{move} = \min[\exp(-\Delta U / kT); 1]$ :

The acceptance probability of insertion is :

$$p_{ins} = \min\left[\frac{V}{N+1} \exp\left(\frac{\mu - \Delta U}{kT}\right); 1\right] \quad (1)$$

while for the particle deletion :

$$p_{del} = \min \left[ \frac{N}{V} \exp\left(\frac{\Delta U - \mu}{kT}\right); 1 \right] \quad (2)$$

where  $\Delta U = U_{new} - U_{old}$  is the potential energy change between the new (trial) and old (current) configuration. The acceptance probability of any new configuration is independent of configurations older than the current. A detailed presentation of the method is described elsewhere (Nicholson & Parsonage, 1982; Allen & Tildesley, 1987). The presence of the adsorbing surface requires the application of periodic boundary conditions depending on the pore geometry. For slit and cylindrical models boundary conditions are applied in the directions other than the width and the diameter respectively, while in the case of the carbon cones periodic conditions are applied in all directions. In our case, each simulation box contains an isolated pore model (slit, cylinder or cone) and statistics are studied only inside the structures (external surface is omitted). In the case of periodic structures (slits and tubes), the size of the box can be varied in order to ensure that at least 500 particles remain in the simulation box at each pressure. The simulations run typically for  $7 \times 10^6$  configurations, while statistics are not collected over the first  $3 \times 10^6$  configurations to assure adequate convergence of the simulation.

## 2.1 Simulation models

### 2.1.1 Gas molecules representation - adsorbate - adsorbate interactions

#### (a) Hydrogen molecule

In general the interaction potential of two particles (i and j) is given by Lennard-Jones (LJ):

$$u_{ij} = 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (3)$$

where  $\epsilon_{ij}$  and  $\sigma_{ij}$  are the characteristic energy and collision diameter of those particles and  $r_{ij}$  is the distance between them. Hydrogen can be treated either as one center (spherical) or as a two center (linear) particle. The parameters for the spherical representation are  $\epsilon_{HH}/k_B = 36.7$  K and  $\sigma_{HH} = 0.2958$  nm (Darkrim & Levesque, 1998) and for the two-site model are  $\epsilon_{HH}/k_B = 12.5$  K and  $\sigma_{HH} = 0.259$  nm, while the H-H distance is assumed to be the actual bond length (0.074 nm) (Cracknell, 2001). Apart from classic (LJ) interaction,  $H_2$  reveals also a quantum behaviour that can contribute significantly to adsorption especially under confinement and/or low temperatures (Sese, 1995; Wang & Johnson, 1999; Tanaka et al., 2005). Thus a correction term is added to the LJ potential and interactions are calculated by the so called Feynmann and Hibbs quantum corrected expression :

$$u_{ij}|_{FH} = u_{ij} + \left( \frac{\beta \hbar^2}{24 \mu_m} \right) \nabla^2 (u_{ij}) \quad (4)$$

where  $\beta = (k_b T)^{-1}$ ,  $\hbar = h/2\pi$  ( $h$  is the Plank constant),  $\mu_m$  is the reduced mass of a pair of interacting molecules ( $\mu_m = m/2$ ) and  $u_{ij}$  is the energy of the pairwise LJ interaction.

#### (b) Carbon dioxide molecule

$CO_2$  is modeled as a three charged center LJ molecule with  $\epsilon_{OO}/k_B = 80.507$  K,  $\sigma_{OO} = 0.3033$  nm,  $\epsilon_{CC}/k_B = 28.129$  K,  $\sigma_{CC} = 0.2757$  nm (Harris & Yung, 1995). The O-O and C-O distances are 0.2298 nm and 0.1149 nm respectively. The model was obtained after suitable optimization

of literature LJ parameters (Murthy et al., 1983), and has the advantage of predicting fairly well the phase coexistence curve and the critical properties of the fluid.

The intermolecular potential  $u_{\text{CO}_2-\text{CO}_2}$  is assumed to be a sum of the interatomic potentials between the atoms of the interacting molecules, plus the electrostatic interactions due to  $\text{CO}_2$  quadrupole moment with point partial charges  $q_1 = q_3 = -0.3256e$  and  $q_2 = +0.6512e$  (Eq.5):

$$u_{\text{CO}_2-\text{CO}_2} = \sum_{i=1}^3 \sum_{j=1}^3 \left( u_{ij} + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right) \quad (5)$$

where  $\epsilon_0$  is the permittivity of vacuum. The indices  $i$  ( $j$ ) refer to the sites of the first (second) interacting molecules. All cross interaction potential parameters between two sites, are calculated according to the Lorentz-Berthelot rules ( $\sigma_{ij} = (\sigma_{ii} + \sigma_{jj})/2$ ,  $\epsilon_{ij} = (\epsilon_{ii}\epsilon_{jj})^{1/2}$ ).

### (c) Nitrogen molecule

$\text{N}_2$  is modelled as a two center LJ molecule (Kuchta & Etters, 1987). The parameters used are  $\epsilon_{\text{NN}}/k_B = 37.8$  K and  $\sigma_{\text{NN}} = 0.3318$  nm, with the two centers separated by 0.1094 nm. The molecule has a quadrupole moment with the four point charges  $q_1 = q_4 = +0.373e$  and  $q_2 = q_3 = -0.373e$  placed along the molecular axis at positions 0.0847 nm and 0.1044 nm from the center respectively. The intermolecular interactions were calculated according to:

$$u_{\text{N}_2-\text{N}_2} = \sum_{j=1}^2 \sum_{i=1}^2 u_{ij} + \sum_{k=1}^4 \sum_{l=1}^4 \frac{q_k q_l}{4\pi\epsilon_0 r_{kl}} \quad (6)$$

where  $i$  and  $k$  or  $j$  and  $l$  refer to the LJ and charge centers of the first or second interacting molecules.

## 2.1.2 Carbon structures – adsorbate / adsorbent interactions

### (a) Slit-shaped pores

Due to the layered graphitic structure, the majority of porous carbons possess slit like pores. Graphitic surfaces can be simulated by stacked planes of LJ carbon atoms separated by  $\Delta = 0.335$  nm and having a number density  $\rho_w = 114$  atoms/nm<sup>3</sup> (figure1). Assuming the classic Lennard - Jones potential for the fluid - wall atom interactions and integrating over all atoms of all graphite planes, the '10-4-3' potential of Steele is deduced (Steele, 1974):

$$u_w(r_z) = 2\pi\rho_w\epsilon_{\alpha\beta}\sigma_{\alpha\beta}^2\Delta \left[ \frac{2}{5} \left( \frac{\sigma_{\alpha\beta}}{r_z} \right)^{10} - \left( \frac{\sigma_{\alpha\beta}}{r_z} \right)^4 - \frac{\sigma_{\alpha\beta}^4}{3\Delta(0.61\Delta + r_z)^3} \right] \quad (7)$$

where  $r_z$  is the distance between a Lennard - Jones site of an adsorbate molecule and the solid surface. The potential parameters of the solid surface are  $\epsilon_{\text{ss}}/k_B = 28.0$  K and  $\sigma_{\text{ss}} = 0.34$  nm, while all the cross interaction potential parameters between different sites are calculated according to the Lorentz-Berthelot rules. The same  $\epsilon_{\text{ss}}$  and  $\sigma_{\text{ss}}$  parameters are used for the cylindrical and the conical carbon structures.

The overall potential energy  $U_w$  due to the walls inside a slit-like pore is calculated by the sum of the interactions between the adsorbate and both pore walls:

$$U_w = u_w(r_z) + u_w(H-r_z) \quad (8)$$

where  $H$  is the distance between the pore walls (physical width).

For interacting with hydrogen slits the F-H quadratic term has to be added to the above expression. In this case this term is (Tanaka et al., 2005):

$$u_{qu}(r_z) = 176\pi\rho_s\varepsilon_{\alpha\beta} \frac{\beta\hbar^2}{24\mu_m} \sum_{j=0}^2 \left[ \frac{1}{2} \left( \frac{\sigma_{\alpha\beta}}{r_z + j\Delta} \right)^{12} - \frac{1}{4.4} \left( \frac{\sigma_{\alpha\beta}}{r_z + j\Delta} \right)^6 \right] \quad (9)$$

so that the total solid-fluid potential,  $u_{H-w}$ , is given by:

$$u_{H-w}(r_z) = u_{10-4-3}(r_z) + u_{qu}(r_z) \quad (10)$$

where  $r_z$  is the distance between the LJ site on the adsorbent and the plane of carbon atoms,  $\mu_m$  is the reduced mass and  $\rho_s=38.19 \text{ nm}^{-2}$  the surface density of a single graphite layer.

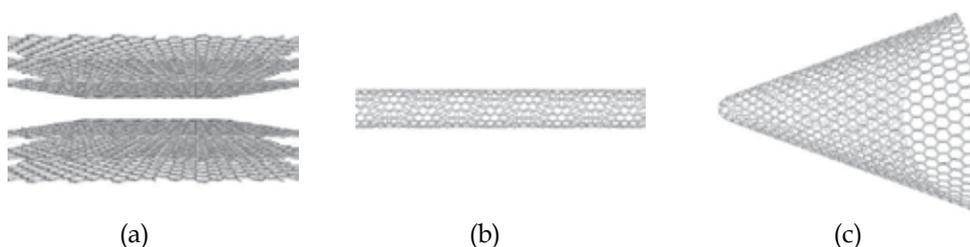


Fig. 1. The carbon pore models a) slit-shaped, b) nanotube and c) cone

### (b) Carbon nanotubes

Carbon nanotubes (CNTs) production was first reported by Iijima (Iijima, 1991). A single wall carbon nanotube (SWNT) is generated (Dresselhauss et al., 1996) by rolling up a graphene sheet into a seamless cylinder (Figure 1). CNT structures are represented by the chiral vector, i.e. a pair of indices  $(n,m)$  referred where  $n$  and  $m$  are scalars on the principal unit vectors of the graphene lattice. The diameter of the nanotube is then given by :

$d = \frac{a}{\pi} \sqrt{(n^2 + nm + m^2)}$  where  $a = 0.246 \text{ nm}$ . The fluid-solid interactions are calculated from equation 4, where the reduced mass of a carbon and a hydrogen atom is:  $\mu_m = 7/12$ .

### (c) Carbon cones

Carbon cones (CCs) are one of the newest carbon structures, revealing unique electronic, chemical and mechanical properties (Heiberg-Andersen & Skjeltorp, 2007; Heiberg-Andersen et al., 2008). CCs were first observed in 1994 (Ge & Sattler, 1994), while bulk quantities of conic structures with five different apex angles were produced later in an industrial process (Krishnan et al., 1997). CC models are also produced by cutting out of a graphene sheet one up to five  $60^\circ$  sectors of and joining the dangling bonds. This procedure introduces 1 up to 5 pentagons in the graphene structure. Therefore, the possible apex angle  $\varphi$  can obtain the discrete values of  $112.9^\circ$ ,  $83.6^\circ$ ,  $60.0^\circ$ ,  $38.9^\circ$  and  $19.2^\circ$  (in accordance with the 5 different observed "real" structures) given by the equation  $\sin(\varphi/2) = (2\pi - p\pi/3)$  for  $p = 1, 2, \dots, 5$ , respectively ( $p$  is the number of pentagons). Each of the five cone structures used in this work contains approximately 2000 carbon atoms and is terminated by hydrogen atoms. The solid-fluid interaction potentials are computed again by equation 4.

### 3. Results and discussion

#### 3.1 The pore wall influence

When a molecule is confined within a pore it interacts with the whole structure resulting in a “complex” potential energy field that depends on the pore shape and width. In narrow micropores (<1nm), the gas-solid interactions are extremely strong, due to the overlapping potentials of opposing walls (Figure 2) and a single energy minimum located at the center of the pore exists. For even smaller pore sizes the repulsive contributions of the opposite walls start to interfere, leading to reduced attraction (i.e. reduced potential well depth). As pore size increases, the solid - fluid potential has two shallow minima and interactions at the center of the pore tend to become negligible. A comparison of the potential functions between slit - shaped and cylindrical pore models of the same dimension (similar pore widths) shows that the curvature of the surface results in deeper minima, i.e. stronger interactions for carbon nanotubes than for slit-shaped pores (Figure 2).

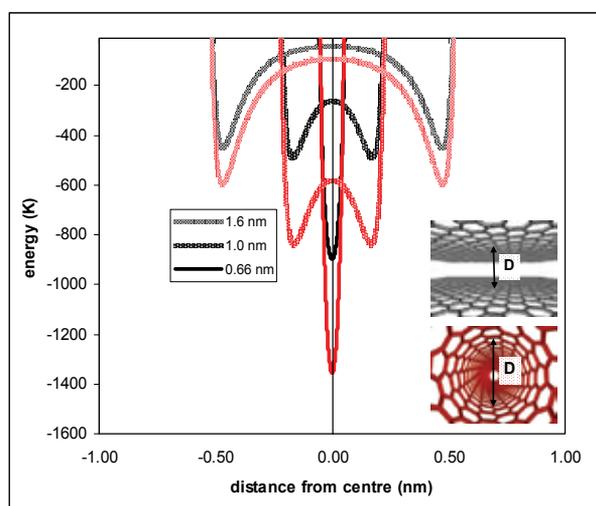


Fig. 2. Hydrogen interaction potential profiles inside carbon slit and nanotube models for three different widths or diameters respectively. Interaction curves are plotted against the distance from the centre.

The interaction potential of a gas molecule inside CCs is similar, however it depends strongly on the distance from the tip since the effective pore width increases linearly with it. Examples of the potential profiles inside CC cavities are presented in figure 3 for regions of high confinement (close to the tip) where interactions are enhanced and obtain a single minimum (a), but also far from the tip (b) (where the diameter is 1nm (Gotzias et al., 2010)).

#### (a) Quantum effects

The contribution of quantum effects on adsorption phenomena depends on the temperature and the density of the fluid inside nanopores. Generally, the quantum behaviour is equivalent to an enlargement of the effective molecule diameter and thus leads to a decrease of the amount adsorbed. A simple criterion regarding the validity of treating a quantum - mechanical system using the classical approach, is based on the de Broglie thermal wavelength  $\lambda$  calculation (Hansen & McDonald, 1990), given by:

$$\Lambda = \frac{h}{\sqrt{2\pi mk_B T}} \quad (11)$$

where  $h$  is the Planck's constant  $m$  is the mass of an atom,  $T$  is the temperature and  $k_B$  is the Boltzmann's constant. Classical approximations are justified, when the ratio  $\Lambda/a$  is much less than unity ( $a \cong \rho^{-1/3}$ , where  $\rho$  is the fluid density). For example, for Ar the ratio is around 0.1 at the triple point ( $T=83.8\text{K}$ ). Hydrogen is much lighter,  $\Lambda/a = 0,94$  at its triple point ( $T=13.8\text{K}$ ) and thus the quantum effects cannot be neglected. Furthermore,  $\Lambda/a$  changes when hydrogen molecules are confined in the narrow micropores, where the strong adsorption potential leads to fluid densification (Liu et al., 1999).

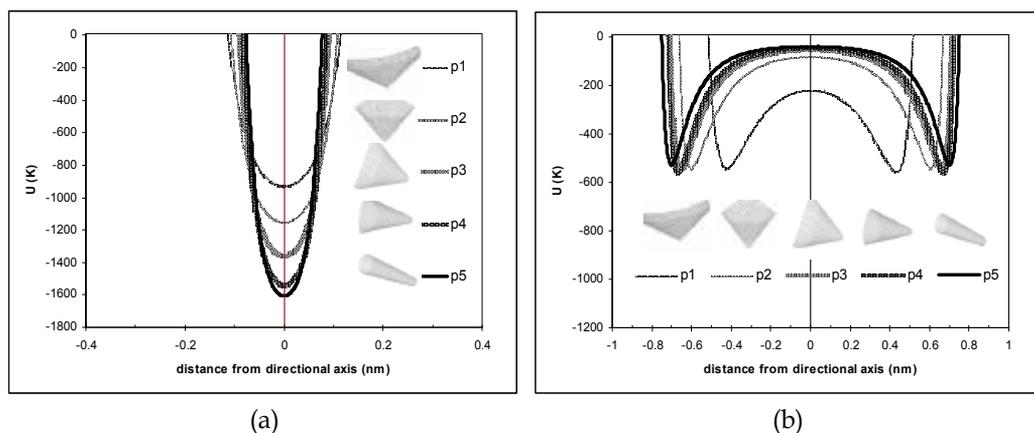


Fig. 3. Interaction profiles inside the five conic cavities. The curves correspond to interaction profiles along the radial distance normal to the cone directional axis a) close to the tip and b) at a circular cone-plane intersection of 1nm diameter

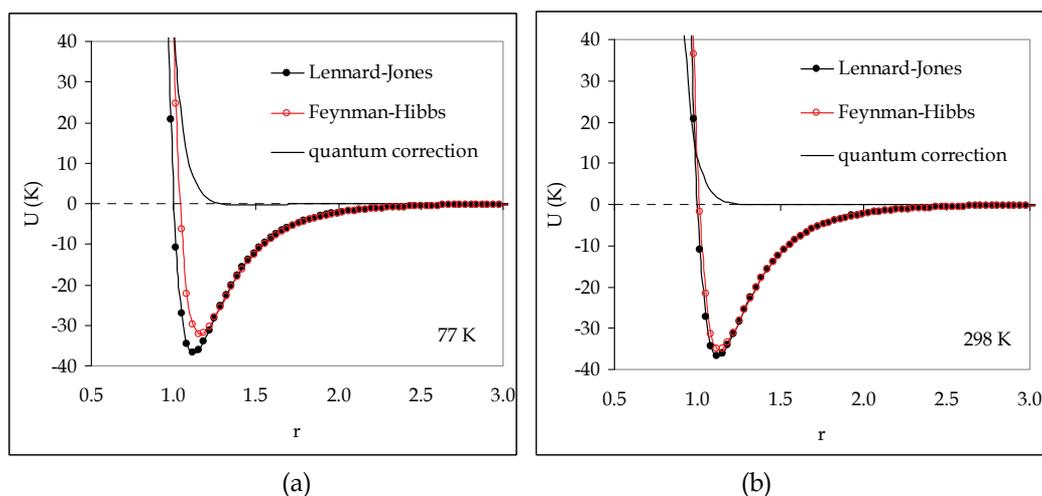


Fig. 4. Classical and Feynman - Hibbs corrected fluid-fluid interaction potential at 77 and 298 K. For comparison the quantum correction has been also plotted.

The results of applying the Feynman – Hibbs effective potential in the fluid – fluid and solid – fluid interactions are demonstrated in figures 4 and 5 at 77 and 298K. Figure 4 presents the total quantum corrected fluid – fluid interaction potential, as well as the classical LJ potential and the quantum contribution calculated by the Feynman – Hibbs method. At low temperature (figure 4a), the quantum contribution is evident, since the total potential reveals weaker interaction between the fluid molecules, than the classical one. On the other hand, at room temperature (figure 4b) the two potentials are very close to each other, however, the classical potential still overestimates slightly the interaction energy.

The differences are more clear when examining the solid – fluid interactions, especially in very small pores. Figure 5 presents the classical and the total quantum corrected solid – fluid interaction potential for selected slit pores. Similarly to the case of fluid – fluid interactions, the classical LJ potential overestimates the interaction energy, while the effect of quantum behaviour depends strongly on the pore size, varying from strong (in smallest pores) to negligible (wide pores).

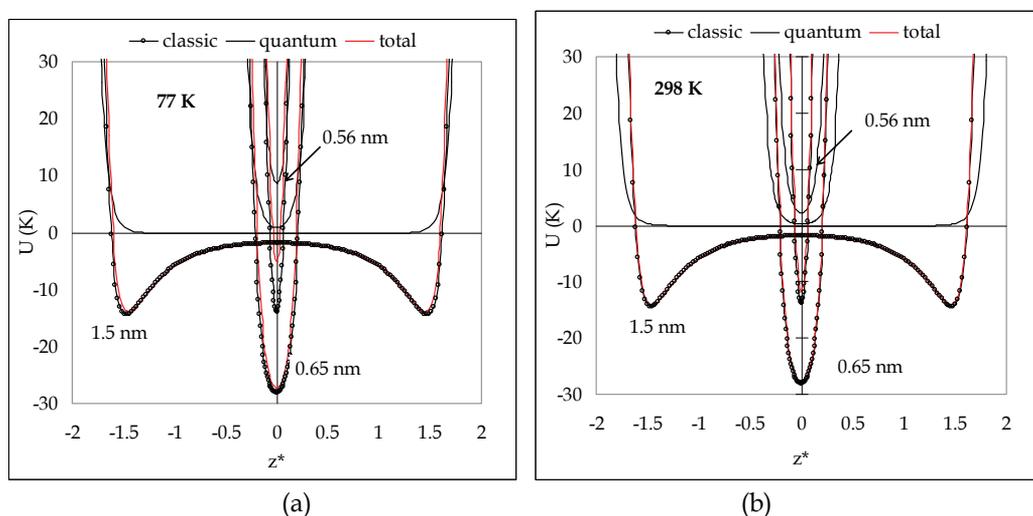


Fig. 5. Classical and Feynman – Hibbs corrected solid-fluid interaction potential at a) 77 and b) 298 K for different slit sizes (the quantum term is also presented).  $z^*$  is the reduced distance from the center ( $z^*=z/\sigma_{HH}$ ).

### 3.2 GCMC simulation results – adsorption isotherms

#### (a) Hydrogen adsorption

GCMC simulations have been carried out for the calculation of the adsorption isotherms of hydrogen in the three carbon structures described above, for pore sizes ranging from about 0.6 to 2.0 nm. The calculations have been performed for discrete pore sizes (per  $\sim 0.1$  nm pore width) for slit-shaped and cylindrical pores (H or D up to 2.0 nm respectively) and for the five possible carbon cones (1-5 pentagons). For comparison, in the cones case, statistics were collected in conic segments of equal volume ( $30 \text{ nm}^3$ ). The simulated isotherms at 77 K and pressures up to 20 bar are presented in figures 6a, 6b (slits and nanotubes respectively) and 7 (cones). For comparison a set of simulated adsorption isotherms at 298 K is also presented.

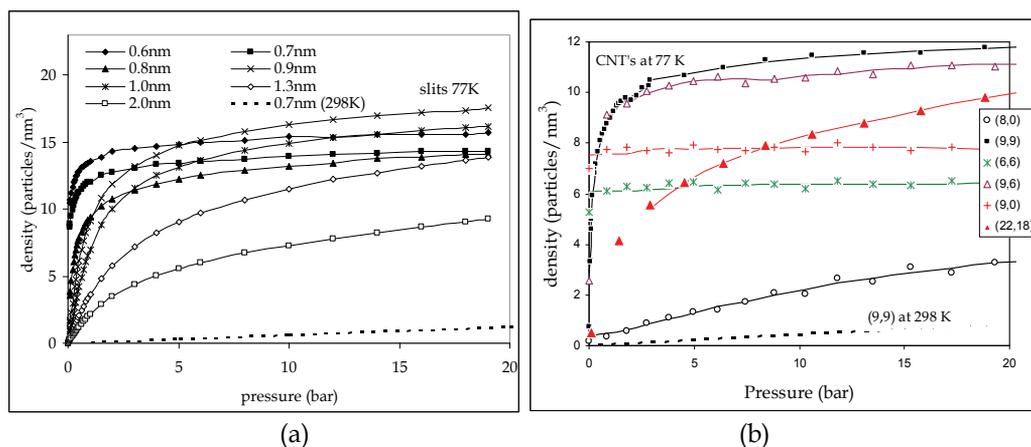


Fig. 6. Calculated adsorption isotherms for hydrogen at 77K in a) slit-shaped pores and b) nanotubes

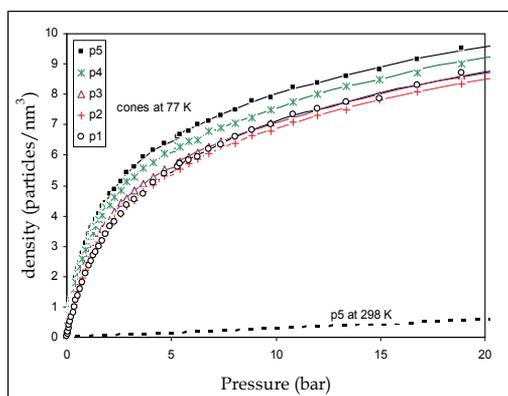


Fig. 7. Hydrogen adsorption densities at 77K corresponding to the five conic cavities of 30 nm<sup>3</sup> volume. The isotherm at 298K for the p5 is also shown.

In the pressure range studied at narrow pores (pore width < 1.0 nm for the slit - shaped and cylinders), the isotherms are of Langmuir type (according to IUPAC classification). Adsorption is particularly enhanced at low pressures, due to the overlapping solid-fluid interaction potential. The very small pores (<0.9nm) are completely filled with hydrogen even for relatively low pressures ( $\approx 1$  bar), and no increase of the amount adsorbed is observed at higher pressures, due to pore volume limitations. As the pore sizes increase the wall interactions grow weaker affecting merely the molecules close to the pore walls, and adsorption is mainly determined by the available pore volume. Therefore isotherms tend to become of Henry type (i.e. straight lines). Similar conclusions are deduced from figure 6b, regarding carbon nanotubes. The basic difference compared to the slit-shaped pores is the much lower density values observed in the fine pores. This can be attributed to the loss of one degree of freedom of the molecules when confined in cylinders, i.e. the system is actually 2D in slits and 1D in cylinders. The nanotubes illustrated in Figure 6b have comparable pore size (diameter) to the slits of figure 6a ((8,0)  $\sim 0.62$ nm, (9,0)  $\sim 0.7$ nm, (6,6)  $\sim 0.8$ nm, (9,6)  $\sim 1.02$ nm, (9,9)  $\sim 1.22$ nm and (22,18)  $\sim 2.7$ nm).

In carbon cones, the differences are less emphasized compared to the slit and cylindrical geometries. The cone heterogeneity, resulting mainly from the curvature gradient along the surface, affects the adsorption process at low pressures where adsorption on the very high energy sites (close to the tip) occurs. After filling these sites the cones reveal a constantly increasing isotherm curve. At low pressures the more "narrow" cones (more pentagons on the tip) have increased adsorption capacity due to stronger confinement, in accordance with the interaction potential profiles of Figure 3. It is however interesting to note that for pressures above 4 bar the capacity of p1 is slightly higher than those of p2 or even p3. This should be expected since at this pressure range adsorption occurs at the core of the pore. There the interaction in p1 is stronger since it has a lower local energetic minimum (Fig. 3b) as the distance from the tip (height) is shorter than the other cones.

### (b) Carbon dioxide adsorption

Carbon dioxide adsorption in nanopores shows many differences from hydrogen. The main reason for this is that while  $H_2$  is supercritical at 77K the critical temperature of carbon dioxide is 31.25°C (304.4K), i.e. high enough to be liquefied even at room temperature. The shape of the isotherms changes gradually, revealing capillary condensation at pressures far below the vapour pressure. The condensation pressure and consequently the final form of the adsorption isotherm, is determined by the pore size and geometry, and also by the fluid-fluid and fluid-solid interaction. In this respect, the carbon dioxide adsorption isotherms in slit-shaped and cylindrical pore models were calculated using the GCMC method at a) 195.5 K the dry ice temperature, below the triple point b) 253 K, where the full (0 - 1) relative pressure range is attained with experimental sorption tests at modest pressures (e.g. up to 20 bar), c) 273 K, a temperature where many adsorption experiments are typically carried out, d) 298 K (room temperature) and e) 308 K, slight above the critical temperature.

The carbon dioxide isotherms obtained by the GCMC simulations for typical slit - shaped pores of selected sizes ( $H$  up to 2.0 nm) and temperatures are presented in Fig. 8. Detailed pertinent work is reported elsewhere (195.5 and 308K: Samios et al., 1997 ; Samios et al., 2000, 253 and 298 K Konstantakou et al., 2007a, Konstantakou et al., 2007b, 195.5, 253 and 273 K Konstantakou et al., 2010). As expected, higher adsorption capacities are observed as temperature decreases. The adsorption mechanism can be divided into three regions: initially at small pressures, micropore filling occurs, where the gas molecules occupy the whole pore space adjacent to the walls, and adsorption is almost entirely controlled by the solid - fluid interactions; at higher pressures, the interaction is getting relatively weaker resulting to a reduced isotherm slope and multiple adsorbed layers are developed in larger pores. The capillary condensation phenomenon takes place abruptly with an almost vertical transition from the gas-like to the liquid-like phase, and the condensation pressure is unique for each pore size, geometry and temperature applied. For this reason such  $CO_2$  computed isotherms can be used for deducing the pore size distribution of porous solids.

Depending on the pore size and the temperature the isotherm may not display all three regions. For example, in very narrow pores (e.g. 0.8 nm) instead of multilayer adsorption/capillary condensation, due to confinement the prevailing mechanism is micropore filling giving rise to Langmuir type isotherms. As the pore size increases (e.g. 1.2 nm) more pore space is available but the wall attraction is weaker and pore filling occurs at higher pressures. For larger pores the isotherm is transformed to type IV. The observed sudden uptake corresponds to the capillary condensation transition. The pores are filled with the liquid - like phase and no further increase in adsorption is observed with pressure.

For wide enough pores, where the wall attraction is almost negligible the fluid density value is approaching the bulk one. Considering the cylindrical geometry, carbon dioxide behavior appears to be rather similar to the slits, except from the smoother densification process. In addition, much lower density values are calculated for the very narrow pores (0.65 nm) contrary to the slits (0.6 nm). For a cylindrical pore, reduction in pore diameter will lead to one-dimensional behavior. As a consequence, molecules cannot attain the most energetically favorable orientation inside the pore and thus pack efficiently, as they are forced to align in a way almost parallel to the pore axis.

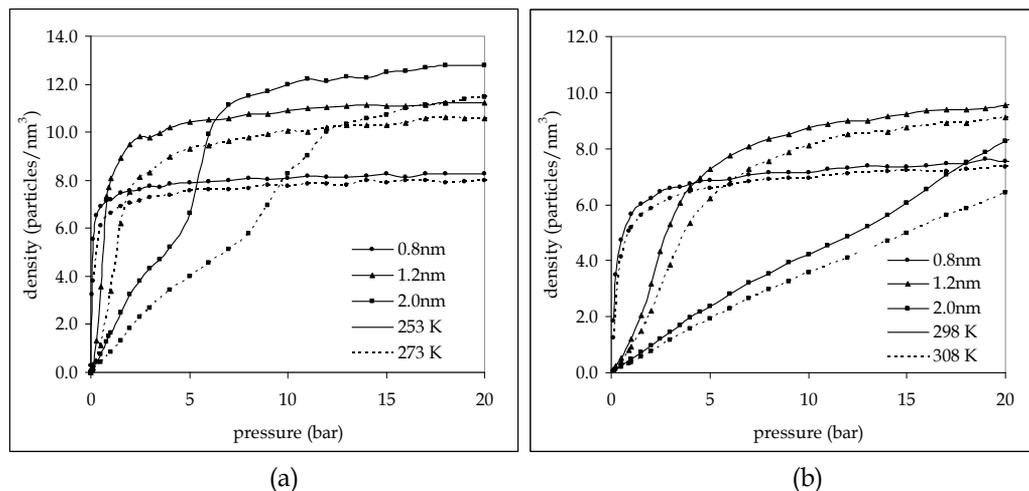


Fig. 8. Carbon dioxide adsorption densities for selected slit pores and temperatures.

In general, the gas-liquid phase transition in pores (capillary condensation) is a very interesting feature, which is observed when condensable gases are confined in small cavities. The phenomenon has been extensively studied in pores of various geometries both experimentally and theoretically (Evans & Tarazona, 1984 ; Ravikovitch et al., 2001 ; Nilson & Griffiths, 1999 ; Neimark et al., 2003). Capillary condensation takes place at pressures below the bulk condensation pressure at a given temperature. The relative pressure where pore condensation appears depends on the liquid interfacial tension, the strength of the attractive interactions among the fluid and pore walls, the pore geometry and the pore size. Moreover, starting from such filled pores and decreasing the pressure (desorption), a hysteresis loop often appears and evaporation occurs at lower pressures than those observed for condensation. For very small pore sizes the condensation-evaporation steps are almost completely reversible, while in larger pores different types of hysteresis loops are observed, depending mainly on the shape and relative dimensions of the pores. In the hysteresis loop region the GCMC fails to adequately describe the gas-liquid coexistence. A direct calculation of the phase co-existence in this metastability region is achieved by using variations of the Gibbs ensemble (Panagiotopoulos, 1987; McGrother & Gubbins, 1999 ; Neimark & Vishnyakov, 2005).

### (c) Nitrogen adsorption

$N_2$  adsorption isotherms at 77 K have been calculated for slit-shaped pores. Two types of adsorption behavior are observed similar to  $CO_2$  case. For small micropores (0.8-1.2nm)

(figure 9a,b) a single layer is formed and pore filling occurs abruptly at low pressures ( $<10^{-4}$  bar), while further pressure increase does not change the amount adsorbed. As pore sizes increase the isotherm gradually changes to type IV. In small pressures, only a few molecules are occupying the sites near the pore walls. As the pressure increases, additional layers are adsorbed since finally the condensation pressure is reached and phase transition occurs. It must be also pointed out that pores smaller than 0.6nm are inaccessible to nitrogen molecules, due to the overlapping repulsive parts of the opposing wall potentials.

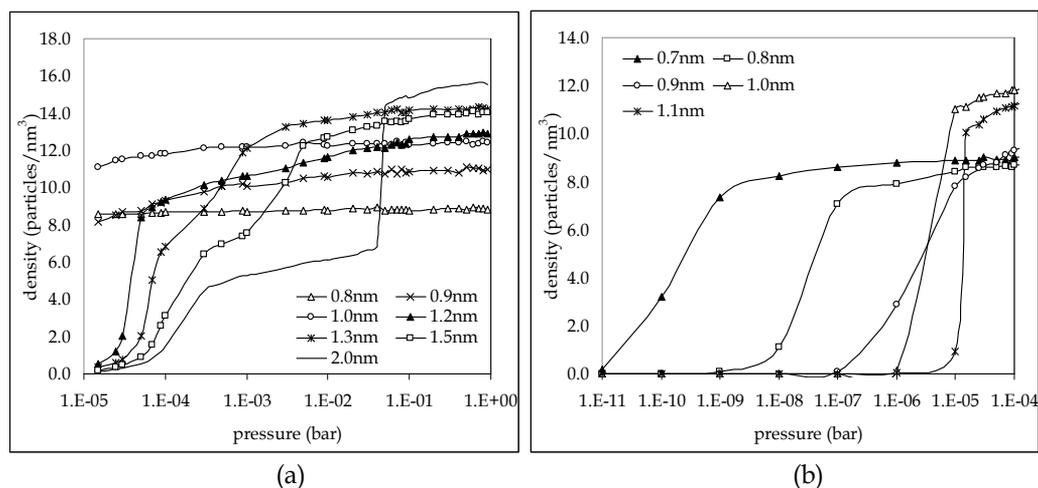


Fig. 9. Nitrogen adsorption densities at 77K for selected slit-shaped pore widths a) up to the vapour pressure  $P_0=1$  bar and b) for very low pressures.

N<sub>2</sub> adsorption isotherms at 77 K are commonly employed for pore size distributions determination of carbonaceous materials. Similar to the CO<sub>2</sub> behavior, nitrogen exhibits capillary condensation in much lower pressures than the vapor pressure, while the step in the isotherm occurs in different values of  $P/P_0$  depending on the pore size. However, for pore widths between 0.7 and 1.0nm, the almost straight horizontal lines appearing for about the entire pressure range that is usually measured experimentally ( $10^{-5}$ -1 bar), imply that the method is not sensitive enough for small micropores, when based solely on N<sub>2</sub> data at 77 K.

### 3.3 GCMC simulation results – density profiles

GCMC simulations provide valuable information regarding the molecular packing and the adsorbed layers formed inside the individual pores at different pressures. The structure of the adsorbate in the pore is reflected by the local density profiles, namely the local number density at a distance  $z$  from the pore center ( $z=0$  at the pore center). The corresponding local orientation of the molecules can also be obtained, as the ensemble average of the directional cosines of the fluid molecules in the cavity. Computationally such density and orientation profiles are obtained by dividing the pore into uniform segments and keeping record of the number of times a molecule is encountered inside each segment during the simulation.

In the case of H<sub>2</sub> the energetically favourable molecular configuration is rather simple, since the molecules are assumed to be chargeless and the fluid-fluid interactions do not have a prominent effect. Figure 10 presents the H<sub>2</sub> density profiles at 77 K in selected slit pores. The very fine pores are completely filled with one layer in their center. As the pore size

increases, the potential starts to reveal two minima and as a result two layers adjacent to the pore walls develop. No other dense layer is formed near the pore center, which is an evidence of the weak type of the fluid – fluid interactions.

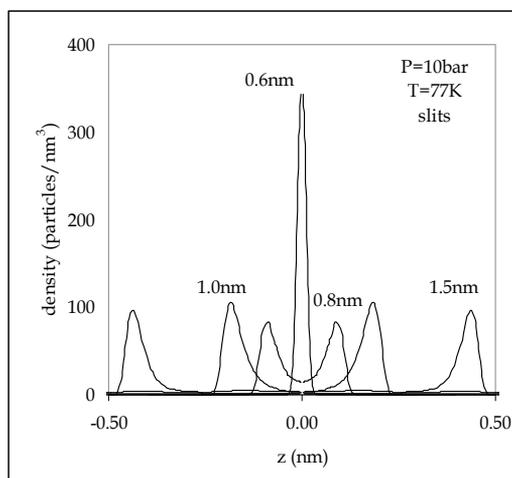


Fig. 10. Hydrogen local density profiles for different pore widths at 77K

Detailed CO<sub>2</sub> density profiles across the pore have been computed for the whole range of micropore widths (from 0.6 to 2.0 nm, in steps of 0.1 nm). The average fluid density in the micropores as a function of the pore width ( $H/2$ ) is presented in Figure 11 for two pressures.

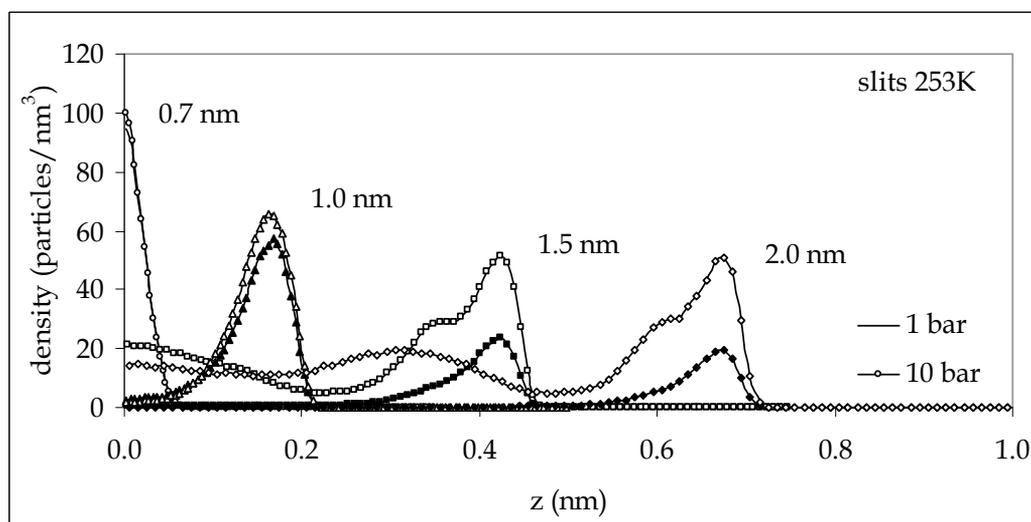


Fig. 11. CO<sub>2</sub> local density profiles for four pore sizes and two selected pressure levels at 253K (filled symbols 1 bar, open symbols 10 bar)

The density profiles reveal the formation of mono and multilayer adsorption films prior to capillary condensation. At low pressures the layer formation follows the same procedure as hydrogen, i.e. one dense layer in the center of the pore that is gradually transformed into

two lower density layers as the pore size increases. At higher pressures further densification of the layers in the smaller pores occurs, while new layers appear in wider pores. From the local orientation profiles it is concluded that the CO<sub>2</sub> molecules constituting the wall layers, tend to lie flat against the surface and this energetically favourable configuration (all three atoms of the molecule in the potential minimum) leads to high average densities in the pore, even at relatively low loadings. For higher pressures, a new distinct sublayer is developed over the primary layer. In the layer closer to the wall the molecules are lying parallel to it and in the second the molecules are oriented mostly normal to the pore wall. The presence of these two differently structured sublayers can be attributed to the quadrupole-quadrupole interactions, which have been reported to contribute essentially to the stability of such a T-like configuration of molecules. The role of the enhanced quadrupole-quadrupole interactions between the carbon dioxide molecules is crucial for sustaining these highly structured configurations. Once the quadrupole moment is eliminated from the simulations, the packing efficiency is lost. A detailed description concerning the local density and orientation profiles of CO<sub>2</sub> in carbon nanopores is given in previous works (Samios et al., 2000 ; Samios et al., 1997).

In Figure 12 the local density profiles of a slit shaped pore with width  $H=2.0$  nm are compared for different temperatures (253 K, 273 K, 298 K and 308 K). It is evident that as temperature increases the wall layers become less dense and those near the pore center less pronounced. Moreover, the sublayer with molecule orientation normal to the wall i.e. the second layer, grows significantly and becomes more structured at lower temperatures. This result is consistent with the expected effect of temperature on the effect of the quadrupole-quadrupole interactions. For the cylindrical geometry the packing behavior of CO<sub>2</sub> is found rather similar. The layers are formed in the same distances from the walls, comparing pores of similar width ( $H \sim D$ ). The main differences marked, concern the height and the width of the layers. In very narrow cylinders, the molecules cannot attain the most favorable orientation, as they are forced to align in a way almost parallel to the pore axis. Also, the sublayers in larger pores are more pronounced, even at high temperatures (298 K).

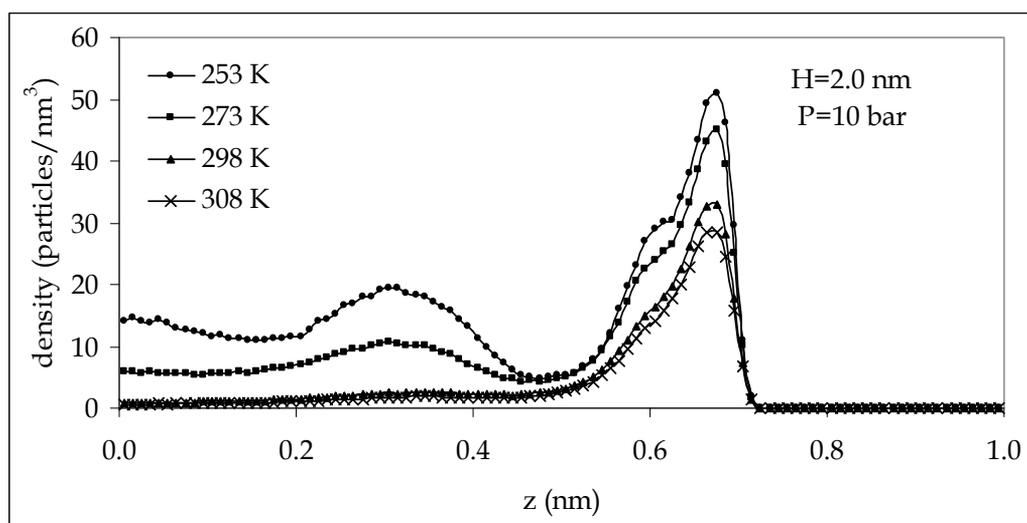


Fig. 12. Temperature effect on the CO<sub>2</sub> local density profiles for 2.0nm slit-shaped pore

### **3.4 Application of the simulated adsorption isotherms to the characterization of carbon porous materials – The pore size distribution (PSD) method**

Gas sorption is considered as a well established method that can provide information on the pore size, the pore network connectivity, and other structural parameters of microporous and mesoporous materials (Kaneko, 1994; Yortsos, 1999). A large number of thermodynamic models of adsorption are commonly used to explain the features of the experimentally obtained isotherms of various gases and vapors. Some of them (Dubinin-Radushkevich, Dubinin-Astakhov and Dubinin-Stoeckli) use the Dubinin theory of pore filling to calculate the structural properties of the material (Greg & Sing, 1982), while others are modified t-curve methods implemented in micropores (MP and Horvarth - Kawazoe method) (Nicholson, 1994 ; Nicholson, 1996 ; Stoeckli et al., 2000). However, the results differ each time the calculation takes place, depending on the model used (Russell & LeVan, 1994 ; Kruk et al., 1998 ; Valladares et al., 1998).

The pore structure of a material is usually described in terms of the pore size distribution (PSD), namely the distribution of pore volume with respect to its size. Over the past 30 years several methods have been developed in this direction. The basic concept relies on the general dependence of the filling pressure and the pore width to the gas amount adsorbed. Since microscopic models can reproduce such a behavior adequately, the Density Functional Theory (Ravikovitch et al., 2006 ; Nguyen & Bhatia, 2004 ; Jagiello & Thommes, 2004) and the Monte Carlo simulation (Do & Do, 2005 ; Nguyen et al., 2005 ; Shao et al., 2004 ; Samios et al., 1997 ; Samios et al., 2000 ; Konstantakou et al., 2007a, Konstantakou et al., 2007b) are the most accepted methodologies. A review on both methods is given by (Do & Do, 2003). Regardless the theoretical approach, the probe molecule usually used for the determination of PSDs in porous carbons is nitrogen at 77 K. The experimental procedure is typical and relatively easy to perform. Additionally, the saturation vapor pressure of nitrogen allows for the accurate measurement of relative pressures over a wide range (commonly  $10^{-5}$  - 0.995). Nitrogen is considered reasonably inert, inexpensive and in abundance, and also its sorption characteristics are well studied in literature. However, many microporous solids have pores with dimensions comparable to the size of nitrogen molecules, prohibiting the convenient diffusion, increasing the equilibration time and leading to significant under estimation of the adsorption isotherm (Rodriguez-Reinoso et al., 1988). Such diffusional limitations could influence adsorption especially in ultra-micropores (<0.7 nm).

Adsorption measurements at higher temperatures (near room temperature) and modest pressures (up to 20 bar) represent a more convenient alternative in terms of both experimental time and resolution (Garrido et al., 1987 ; Sweatman & Quirke, 2001) since the higher temperature facilitates molecules to access the narrower pores. For instance, a comparison between the PSDs calculated for the same material (activated carbon AX-21) using as probe molecules carbon dioxide at 293,1 K and nitrogen at 77 K, revealed that carbon dioxide can detect different pore sizes and in particular smaller than nitrogen (Scaife et al., 2000). Similarly, hydrogen is considered an excellent probe for very fine pores due to its small size, while at 77 K it is far above its supercritical temperature ensuring fast equilibration kinetics. Pertinent adsorption measurements have been used lately for the determination of PSDs (Jagiello & Thommes, 2004; Konstantakou et al., 2007a; Konstantakou et al., 2007b).

#### **(a) Description of the method**

The procedure for the determination of the optimal PSD of microporous carbonaceous materials requires two sets of data. The first set contains the experimental results of one or

more measured isotherms, depending on the number of gases used and / or the different temperatures applied. The other set (kernel) is composed of the corresponding (i.e. for the gases and temperatures used experimentally) simulated isotherms for different pore sizes. In order to derive the local isotherms (amount adsorbed for each pore size) the procedure described must be followed, i.e. accurate construction of pore models, precise representation of the probe molecule and appropriate selection of potential models. The method is based on the assumption that the experimental isotherm consists of a number of individual local (single pore) isotherms, each multiplied by a scaling factor (i.e. relative volume that each pore size occupies in the material). The adsorption integral has the following expression.

$$b(p) = \int_{H_{\min}}^{H_{\max}} A(H, p) \cdot x(H) dH \quad (12)$$

where  $H$  and  $p$  are the pore width and the pressure step respectively.  $A(H, p)$  represents the adsorption kernel,  $b(p)$  denotes the experimental isotherm and  $x(H)$  is the unknown pore size distribution that covers a realistic range of pore sizes ( $H_{\min}$ ,  $H_{\max}$ ). Actually this is a linear system of  $m$  equations and  $n$  unknowns (i.e.  $b=Ax$ ) which can commonly be solved by minimizing the residual

$$\text{Minimize } \frac{1}{2} \|b - Ax\|^2 \text{ in respect to constraints} \quad (13)$$

The solution refers to a constrained least squares problem. The constraints are simple as we expect only non negative solutions and that the cumulative pore size distribution is equal to unity. Further constraints on the smoothness of  $x$  in order to force physically sound or appealing solutions, can also be imposed (see Konstantakou et al., 2007a; Konstantakou et al., 2007b and references therein). For this work the E04NCF routine from the NAG library (Gill et al., 1984; Stoer, 1971) has been implemented in order to perform the minimization.

### (b) Application example

The technique is applied in the case of the KOH activated carbon AX-21, a microporous material with large surface area (made by Amoco Co. and kindly provided by S. R. Tenisson, MAST Carbon). In order to assure the reliability of the deduced PSD, several experimental isotherms were used, based on the argument that different adsorption isotherms in terms of temperatures and/or gases can probe different pore sizes. Therefore, instead of using  $N_2$  at 77 K, high temperature  $CO_2$  and  $H_2$  isotherms at 77K have been considered. The measurements of the adsorption isotherms of  $CO_2$  (253 and 298) and  $H_2$  (77 K) were carried out in a pressure range 0–20 bar on the Intelligent Gravimetric Analyzer (IGA, Hiden Analytical Ltd.). Consequently, 25 GCMC adsorption isotherms were calculated overall, for each gas and temperature in the pore range  $H=0.6 - 3.0$  nm (the pore range was divided in 25 equidistant intervals (pore groups) with 0.1 nm spacing between them). We must point out, that in the simulations the slit-shaped pore model was used, since it approximates better the sample structure.

In the first stage, three different PSDs were deduced (Fig. 13a), based on the individual  $CO_2$  at 253 K,  $CO_2$  at 298 K and  $H_2$  at 77 K adsorption data and the respective experimental isotherms. As expected, all the distributions produced, are presenting dissimilar shape, as different probe molecules detect different pore groups.

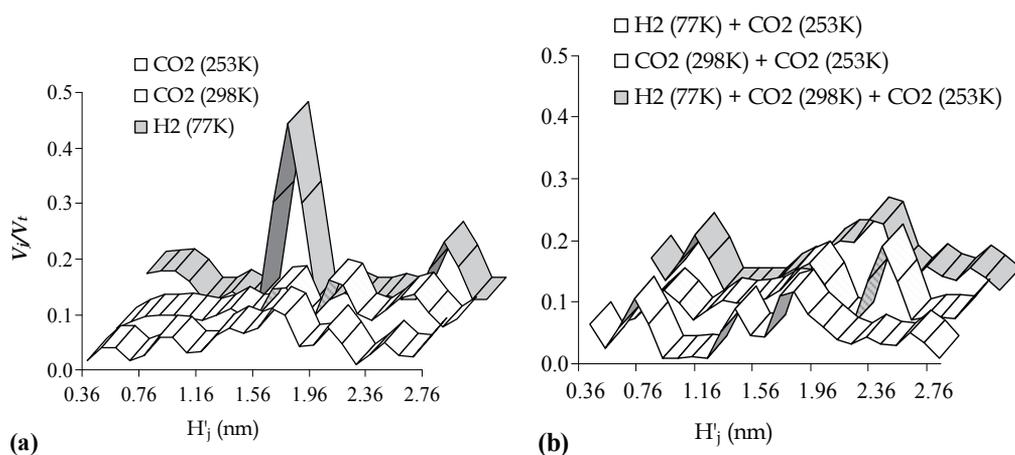


Fig. 13. PSDs based on a) individual adsorption isotherms, b) combinations of adsorption isotherms ( $V_j$ : volume of each class of pores and  $V_t$ : total pore volume).

The PSD obtained from the H<sub>2</sub> data is in marked contrast with that of CO<sub>2</sub>, as it practically reveals 3 classes of pores centered at around  $H = 0.7, 1.6$  and  $2.7$  nm. The simulated hydrogen isotherms lose gradually their curvature for pore sizes over  $1.2$  nm, turning to straight lines. In this respect, H<sub>2</sub> sorption isotherms cannot be used for the pore size characterization of samples having pores beyond the ultramicropore region, unless of course adsorption experimental data at much larger pressures are available. On the other hand, the PSDs of CO<sub>2</sub> at both temperatures display a much broader distribution of sizes spread over the entire range studied. Considering that the experimental isotherms were carried out at exactly the same equilibration pressures (up to 20 bar), much lower relative pressures have been measured at 298 K, while the measurement at 253 K contains the full relative pressure range (up to  $P/P_0$  0.94). Consequently, the 298 K isotherm is expected to depict better the contribution of fine pores and the information given for the large pores is minimal. More complete information regarding the large pore region results from the 253 K isotherm.

By simultaneously inverting different adsorption integral equations, after including different combinations of experimental and GCMC data sets (but the same  $f(H)$  function), new PSDs are obtained (Fig.13b). The three "combined" PSDs are similar to each other, covering a long range of pore widths. The reliability of the PSDs is examined by using them in a reverse manner, i.e. for the prediction of the adsorption isotherms. The above approach has been followed for all the calculated PSDs and selected results are presented in Fig. 14. Of course all the PSDs can quite accurately predict their experimental counterparts, nevertheless experimental data on different molecules and/or temperatures cannot be accurately reproduced as presented in figure 14a (example of experimental CO<sub>2</sub> data at 253 K and predictions based on the PSDs of figure 13a). In contrast to the individual, the "combined" PSDs can better reproduce more than one experimental isotherm. As expected, the actual porous system of AX-21 is more accurately depicted by the PSD obtained from the combination of all the data. It should also be mentioned that the combination of the two "low relative pressure" isotherms (H<sub>2</sub> and CO<sub>2</sub> at 298 K) could not reproduce the CO<sub>2</sub> isotherm at 253 K (again because large pore information is missing). A straightforward result concerning the implementation of the technique is that the use of different probes is

essential for the valid characterization of polydisperse porous materials. The reliable PSD should contain complete information of a full relative pressure range isotherm and an accurate description of ultramicropores.

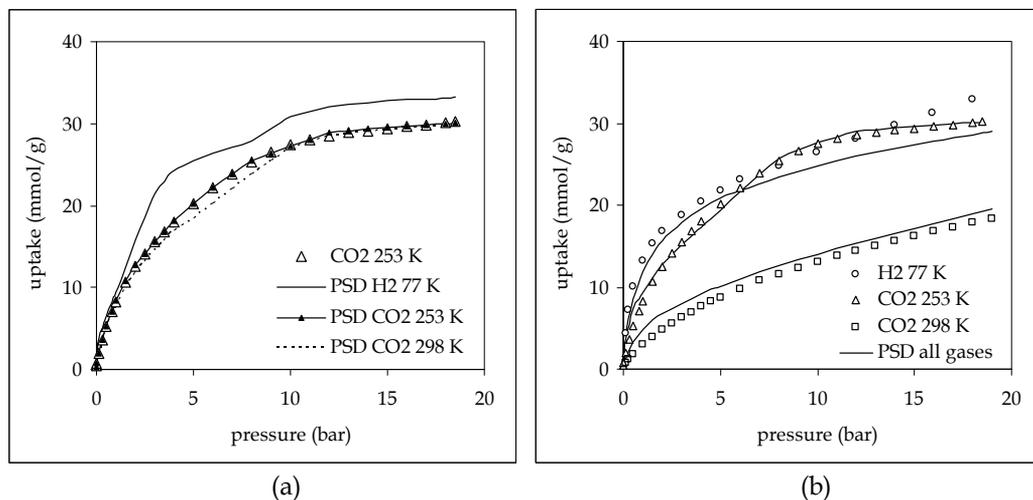


Fig. 14. a) Experimental CO<sub>2</sub> isotherm at 253 K (open triangles) and pertinent predictions based on the individual PSDs of figure 13a. b) Experimental isotherms (open circles: H<sub>2</sub> 77 K, open triangles: CO<sub>2</sub> 253 K, open squares: CO<sub>2</sub> 298 K) on AX-21 and their predictions (lines) based on the combined PSDs using all data

#### 4. Conclusions

The Grand Canonical Monte Carlo method is applied for the study of gas sorption (hydrogen, carbon dioxide and nitrogen) into narrow carbon pores of slit, cylindrical and conical shape. Depending on the gas, the calculations are performed for different temperatures (H<sub>2</sub> at 77 and 298K, CO<sub>2</sub> at 195.5, 253, 273, 298 and 308K and N<sub>2</sub> at 77K). Based on these results an analysis of the distinctive behaviour of each gas molecule when confined in carbon micropores is presented, while the effects of temperature and pore size/geometry is pointed out. In micropores the attractive forces between gas molecules and the surrounding pore walls are intense due to the overlapping solid wall potentials, resulting in single deep potential wells and thus enhanced gas adsorption. In the case of hydrogen, specific attention is given to the significance of employing the proper potentials for the description of the solid-fluid and fluid-fluid interactions. Evidently, quantum effects play an important role in hydrogen adsorption, especially at low temperatures and in the smallest pores.

Additionally, many interesting features are remarked from the calculated adsorption isotherms. Hydrogen sorption is in favour of very fine pores and low temperatures, while for wider pores adsorption capacity depends on the available surface area. In cones the enhanced adsorption is mainly attributed to the tip area, but also to the unique combination of local curvature and confinement. On the other hand, the polar nature of carbon dioxide and nitrogen molecules along with the introduction of wall forces, lead to interesting phase transitions. The simulation results also provide useful insight concerning the packing of the

gas molecules in the individual pores (local density and orientation profiles) as temperature and pressure change.

Finally, the simulated isotherms are used in combination with experimental data, in order to characterize microporous carbons and obtain optimal pore size distributions (PSDs). The adsorption isotherms have been used either individually or in a combined manner in order to deduce the PSDs and their reliability was examined by the ability to predict the experimental adsorption isotherms. Each probe molecule can detect different range of pore sizes. Thus, the combined approach was found to be capable of reproducing more accurately all the available experimental isotherms.

## 5. References

- Alder, B.J. & Wainwright, T.E. (1960). Studies in Molecular Dynamics. II. Behavior of a Small Number of Elastic Spheres. *J. Chem. Phys.*, 33, 6, (November 1960) 1439-51, 0021-9606
- Allen, M.P. & Tildesley, D.J. (1987). *Computer Simulation of Liquids*, Clarendon press, 0198556454, Oxford
- Aukett, P.N.; Quirke, N.; Riddiford, S. & Tennison, S.R. (1992). Methane adsorption on microporous carbons—A comparison of experiment, theory, and simulation. *Carbon*, 30, 6, 913-24, 0008-6223
- Cracknell, R.F. (2001). Molecular simulation of hydrogen adsorption in graphitic nanofibres. *Phys. Chem. Chem. Phys.*, 3, 11, 2091-7, 1463-9076
- Darkrim, F. & Levesque, D.J. (1998). Monte Carlo simulations of hydrogen adsorption in single-walled carbon nanotubes. *Chem. Phys.*, 109, 12, 4981-4, 0301-0104
- Do, D.D. & Do, H.D. (2003). Pore Characterization of Carbonaceous Materials by DFT and GCMC Simulations: A Review. *Adsorption Science and Technology*, 21, 5, (June 2003) 389-423, 0263-6174
- Do, D.D. & Do, H.D. (2005). Comparative adsorption of spherical argon and flexible n-butane in carbon slit pores—a GCMC computer simulation study. *Colloids Surf. A: Physicochem. Eng. Aspects*, 252, 1, (January 2005) 7-20, 0927-7757
- Dresselhauss, M.S.; Dresselhauss, G. & Eklund, P.C. (1996). *Science of fullerenes and carbon nanotubes*, Academic Press, 0122218205, San Diego
- Evans, R. & Tarazona, P. (1984). Theory of Condensation in Narrow Capillaries. *Physical Review Letters*, 52, 7, 557-60, 0031-9007
- Everett D.H. (1972). *Manual of Symbols and Terminology for Physicochemical Quantities and Units*, Appendix II: Definitions, Terminology and Symbols in Colloid and Surface Chemistry. *Pure Appl. Chem.*, 31, 4, 577-638, 0033-4545
- Frenkel, D. & Smit, B. (1996). *Understanding Molecular Simulation, from Algorithms to Applications*, Academic Press, 0122673700, USA
- Garrido, J.; Linares-Solano, A.; Martin-Martinez, J.M.; Molina-Sabio, M.; Rodriguez-Reinoso, F. & Torregosa, R. (1987). Use of nitrogen vs. carbon dioxide in the characterization of activated carbons. *Langmuir*, 3, 1, (January 1987) 76-81, 0743-7463
- Ge, M. & Sattler, K. (1994). Observation of fullerene cones. *Chem. Phys. Lett.*, 220, 3-5, 192-6, 0009-2614
- Gill, P.E.; Murray, W.; Saunders, M.A. & Wright, M.H. (1984). Procedures for Optimization Problems with a Mixture of Bounds and General Linear Constraints. *ACM Trans. Math. Software*, 10, 3, (September 1984) 282-98, 0098-3500

- Gotzias, A.; Heiberg-Andersen, H.; Kainourgiakis, M. & Steriotis, Th.A. (2010). Grand canonical Monte Carlo simulations of hydrogen desorption in carbon cones. *Applied Surface Science*, 256, 17, 5226–31, 0169-4332
- Gregg, S.J. & Sing, K.S.W. (1982). *Adsorption Surface Area and Porosity*, Academic Press, 0123009561, London
- Gregg, S.J. & Sing, K.S.W. (1982). *Adsorption, Surface Area and Porosity*, Academic Press, 0123009561, London
- Hansen, J.P. & McDonald, I.R. (1990). *Theory of Simple Liquids*; Academic Press, 0123705355, London
- Harris, J.G. & Yung, K.H. (1995). Carbon Dioxide's Liquid-Vapor Coexistence Curve And Critical Properties as Predicted by a Simple Molecular Model. *J. Phys. Chem.*, 99, 31, 12021-4, 0022-3654
- Heiberg-Andersen, H. & Skjeltorp, A.T. (2007). Spectra of Conic Carbon Radicals. *J. Math. Chem.*, 42, 4, (November 2007) 707-27, 0259-9791
- Heiberg-Andersen, H.; Skjeltorp, A.T. & Sattler, K.J. (2008). Carbon nanocones: A variety of non-crystalline graphite. *Non-Cryst. Solids* 354, 47-51, (December 2008) 5247-9, 0022-3093
- Iijima, S. (1991). Helical microtubules of graphitic carbon. *Nature*, 354, 6348, (November 1991) 56-7, 0028-0836
- Jagiello, J. & Thommes, M. (2004). Comparison of DFT characterization methods based on N<sub>2</sub>, Ar, CO<sub>2</sub>, and H<sub>2</sub> adsorption applied to carbons with various pore size distributions. *Carbon*, 42, 7, (February 2004) 1227-32, 0008-6223
- Kaneko, K. J. (1994). Determination of pore size and pore size distribution: 1. Adsorbents and catalysts. *Membrane Sci.*, 96, 1-2, (November 1994) 59-89, 0376-7388
- Konstantakou, M.; Samios, S.; Steriotis, Th.A.; Kainourgiakis, M.; Papadopoulos, G.K.; Kikkinides, E.S. & Stubos, A.K. (2007). Determination of Pore Size Distribution in Microporous Carbons Based on CO<sub>2</sub> and H<sub>2</sub> Sorption Data. *Studies in Surface Science and Catalysis*, 160, 543-50, 0167-2991
- Konstantakou, M.; Steriotis, Th.A.; Papadopoulos, G.K.; Kainourgiakis, M.; Kikkinides, E.S. & Stubos, A.K. (2007). Characterization of Nanoporous Carbons by Combining CO<sub>2</sub> and H<sub>2</sub> Sorption Data with the Monte Carlo Simulations. *Applied Surface Science*, 253, 13, 5715-20, 0169-4332
- Konstantakou, M.; Steriotis, Th.A.; Kikkinides, E.S. & Stubos, A.K. (2010). Monte Carlo simulations of CO<sub>2</sub> sorption in nanoporous carbons. *Special Topics & Reviews in Porous Media – An International Journal*, 1, 3, 205-13, 2151-4798
- Krishnan, A.; Dujardin, E.; Treacy, M. M. J.; Hugdahl, J.; Lynam, S.; & Ebbesen, T, W (1997). Graphitic Cones and the nucleation of curved carbon surfaces, *Nature*, 388, (July 1997), 451-454, 0028-0836
- Kruk, M.; Jaroniec, M. & Choma, J. (1998). Comparative analysis of simple and advanced sorption methods for assessment of microporosity in activated carbons. *Carbon*, 36, 10, (October 1998) 1447-58, 0008-6223
- Kuchta, B. & Eters, R.D.(1987). Calculated properties of monolayer and multilayer N<sub>2</sub> on graphite. *Phys. Rev. B*, 36, 6, 3400-06, 0163-1829
- Lastoskie, C.; Gubbins, K.E. & Quirke, N. (1993). Pore size distribution analysis of microporous carbons: a density functional theory approach. *J. Phys. Chem.*, 97, 18, 4786-96, 0022-3654

- Liu, C.; Fan, Y.Y.; Liu, M.; Cong, H.T.; Cheng, H.M. & Dresselhaus, M.S. (1999). Hydrogen Storage in Single-Walled Carbon Nanotubes at Room Temperature. *Science*, 286, 5442, 1127-9, 0036-8075
- Lowell, S. & Shields, J.E. (1991). Powder Surface Area and Porosity, Chapman Hall, 0412396904, London
- Mcgrother, S.C. & Gubbins, K.E. (1999). Constant Pressure Gibbs Ensemble Monte Carlo Simulations of Adsorption into Narrow Pores. *Molecular Physics*, 97, 8, 955-65, 0026-8976
- Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H. & Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.*, 21, 6, (June 1953) 1087-92, 0021-9606
- Murthy, C.S.; O'Shea, S.F. & McDonald, I.R. (1983). Electrostatic interactions in molecular crystals Lattice dynamics of solid nitrogen and carbon dioxide. *Molecular Physics*, 50, 3, 531-541, 0026-8976
- Neimark, A.V.; Ravikovitch, P.I. & Vishnyakov, A. (2003). Bridging Scales from Molecular Simulations to Classical Thermodynamics: Density Functional Theory of Capillary Condensation in Nanopores. *J. Phys.: Condens. Matter*, 15, 3, 347-65, 0953-8984
- Neimark, A.V. & Vishnyakov, A. (2005). A Simulation Method for the Calculation of Chemical Potentials in Small, Inhomogeneous, and Dense Systems. *The Journal of Chemical Physics*, 122, 23, (June 2005) 234108-19, 0021-9606
- Nguyen, T.X. & Bhatia, S.K. (2004). Probing the Pore Wall Structure of Nanoporous Carbons Using Adsorption. *Langmuir*, 20, 9, (March 2004) 3532-5, 0743-7463
- Nguyen, T.X.; Bhatia, S.K. & Nicholson, D. (2005). Prediction of High-Pressure Adsorption Equilibrium of Supercritical Gases Using Density Functional Theory. *Langmuir*, 21, 7, (February 2005) 3187-97, 0743-7463
- Nicholson, D. & Parsonage, N.G. (1982). Computer Simulation and the Statistical Mechanics of Adsorption, Academic Press, 0125180608, London
- Nicholson, D.J. (1994). Simulation study of nitrogen adsorption in parallel-sided micropores with corrugated potential functions. *Chem. Soc., Faraday Trans.*, 90, 1, 181-6, 0956-5000
- Nicholson, D.J. (1996). Using computer simulation to study the properties of molecules in micropores. *Chem. Soc., Faraday Trans.*, 92, 1, 1-10, 0956-5000
- Nilson, R.H. & Griffiths, S.K. (1999). Condensation Pressures in Small Pores: An Analytical Model Based on Density Functional Theory. *Journal of Chemical Physics*, 111, 9, (June 1999) 4281-90, 1089-7690
- Nilson, T.; Nicholson, D. & Kaneko, K. (2003). Temperature Dependence of Micropore Filling of N<sub>2</sub> in Slit-Shaped Carbon Micropores: Experiment and Grand Canonical Monte Carlo Simulation. *Langmuir*, 19, 14, 5700-5707, 0743-7463
- Panagiotopoulos, A.Z. (1987). Adsorption and Capillary Condensation of Fluids in Cylindrical Pores by Monte Carlo Simulation in the Gibbs Ensemble. *Molecular Physics*, 62, 3, 701-19, 0026-8976
- Ravikovitch, P.I.; O'Domhnaill, S.C.; Neimark, A.V.; Schuth, F. & Unger, K.K. (1995). Capillary Hysteresis in Nanopores: Theoretical and Experimental Studies of Nitrogen Adsorption on MCM-41. *Langmuir*, 11, 12, 4765-72, 0743-7463

- Ravikovitch, P.I.; Vishnyakov, A. & Neimark, A.V. (2001). Density Functional Theories and Molecular Simulations of Adsorption and Phase Transitions in Nanopores. *Physical Review E*, 64, 1, 011602-22, 1539-3755
- Ravikovitch, P.I.; Vishnyakov, A.; Neimark, A.V.; Ribeiro Carrott, M.M.L.; Russo, P.A. & Carrott, P.J. (2006). Characterization of Micro-Mesoporous Materials from Nitrogen and Toluene Adsorption: Experiment and Modeling. *Langmuir*, 22, 2, (November 2005) 513-6, 0743-7463
- Rodriguez-Reinoso, F. & Linares-Solano, A. (1988). Microporous Structure of Activated Carbons as Revealed by Adsorption Methods, In: *Chemistry and Physics of Carbon*, vol. 21, P.A. Thrower (Ed.), Marcel Dekker, 978-0-8247-7939-9, New York
- Russell, B.P. & LeVan, M.D. (1994). Pore size distribution of BPL activated carbon determined by different methods. *Carbon*, 32, 5, (January 1994) 845-55, 0008-6223
- Ruthven, D.M. (1984). *Principles of Adsorption and Adsorption Processes*, Wiley - Interscience, 0471866067, New York
- Samios, S.; Stubos, A.K.; Kanellopoulos, N.K.; Cracknell, R.F.; Papadopoulos, G.K. & Nicholson, D. (1997). Determination of Micropore Size Distribution from Grand Canonical Monte Carlo Simulations and Experimental CO<sub>2</sub> Isotherm Data. *Langmuir*, 13, 10 2795-2802, 0743-7463
- Samios, S.; Stubos, A.; Papadopoulos, G.K.; Kanellopoulos, N.K. & Rigas, F. (2000). The Structure of Adsorbed CO<sub>2</sub> in Slitlike Micropores at Low and High Temperature and the Resulting Micropore Size Distribution Based on GCMC Simulations. *J. Colloid Interface Sci.*, 224, 2, 272-90, 0021-9797
- Scaife, S.; Kluson, P. & Quirke, N. (2000). Characterization of Porous Materials by Gas Adsorption: Do Different Molecular Probes Give Different Pore Structures?. *J. Phys. Chem. B*, 104, 2, (December 1999) 313-8, 1089-5647
- Seaton, N.A.; Walton, J.P.R.B. & Quirke, N. (1989). A new analysis method for the determination of the pore size distribution of porous carbons from nitrogen adsorption measurements. *Carbon*, 27, 6, 853-61, 0008-6223
- Sese, L.M. (1995). Feynman-Hibbs potentials and path integrals for quantum Lennard-Jones systems: Theory and Monte Carlo simulations. *Mol. Phys.*, 85, 5, 931-47, 0026-8976
- Shao, X.; Wang, W.; Xue, R. & Shen, Z. (2004). Adsorption of Methane and Hydrogen on Mesocarbon Microbeads by Experiment and Molecular Simulation. *J. Phys. Chem. B*, 108, 9, (February 2004) 2970-8, 1089-5647
- Sosin, K.A. & Quinn, D.F. (1995). Using the high pressure methane isotherm for determination of pore size distribution of carbon adsorbents. *J. Porous Mater.*, 1, 1, 111-19, 1380-2224
- Steele, W.A. (1974). *The Interaction of Gases with Solid Surfaces*, Pergamon, 0080177247, Oxford
- Stoeckli, F.; Guillot, A.; Hugli-Cleary, D. & Slasli, A.M. (2000). Pore size distributions of active carbons assessed by different techniques. *Carbon*, 38, 6, (February 2000) 938-41, 0008-6223
- Stoer, J. (1971). On the Numerical Solution of Constrained Least-Squares Problems. *SIAM J. Numer. Anal.*, 8, 2, (June 1971) 382-411, 0036-1429 (print) 1095-7170 (online)
- Sweatman, M.B. & Quirke, N.J. (2001). Characterization of Porous Materials by Gas Adsorption at Ambient Temperatures and High Pressure. *J. Phys. Chem. B*, 105, 7, (January 2001) 1403-1411, 1089-5647

- Tanaka, H.; Fan, J.; Kanoh, H.; Yudasaka, M.; Iijima, S. & Kaneko, K. (2005). Quantum nature of adsorbed hydrogen on single-wall carbon nanohorns. *Mol. Simul.*, 31, 6-7, 465-74, 0892-7022
- Tanaka, H.; Kanoh, H.; Yudasaka, M.; Iijima, S. & Kaneko, K. (2005). Quantum Effects on Hydrogen Isotope Adsorption on Single-Wall Carbon Nanohorns. *J. Am. Chem. Soc.*, 127, 20, 7511-16, 0002-7863
- Valladares, D. L.; Rodriguez-Reinoso, F. & Zgrablich, G. (1998). Characterization of active carbons: the influence of the method in the determination of the pore size distribution. *Carbon*, 36, 10, (October 1998) 1491-9, 0008-6223
- Wang, Q. & Johnson, J.K. (1999). Computer Simulations of Hydrogen Adsorption on Graphite Nanofibers. *J. Phys. Chem. B*, 103, 2, 277-81, 1089-5647
- Yortsos, Y.C. (1999). Probing Pore Structures by Sorption Isotherms and Mercury Porosimetry, In: *Methods of the Physics of Porous Media, Volume 35 (Experimental Methods in the Physical Sciences, Po-Zen Wong, (Ed.), 69-117, Academic Press, 0124759823, New York*

# Effect of the Repulsive Interactions on the Nucleation and Island Growth: Kinetic Monte Carlo Simulations

Hu Juanmei and Wu Fengmin

*Institute of Condensed Matter Physics, Zhejiang Normal University,  
Jinhua, Zhejiang 321004  
People's Republic of China*

## 1. Introduction

The initial stages of the thin-film epitaxy, such as nucleation and island growth, are important to the synthesis of a wide variety of interfacial materials. The island nucleation process is typically described by the mean-field nucleation theory (Venables J. A., 1973; Venables J. A. et al., 1984), which can be solved to predict the density of stable islands as a function of the deposition rate and the diffusivity of an isolated adatom at low temperatures, such as  $N_x \sim (F/D)^{1/3}$ . Its validity has been investigated experimentally in studies of several homo- and hetero-epitaxial growth systems (Brune H., 1998; Brune H. et al., 1999; Amar J. G. & Family F., 1995; Ratsch C. et al., 1994). However, both experiment and theory indicate that the nucleation theory is no longer valid when the adatom-adatom interactions beyond the nearest-neighbor sites are taken into account at low temperatures (Brune H. et al., 1995; Barth J.V. et al., 2000; Fischer B. et al., 1999; Bogicevic A., et al., 2000).

The long-range interactions between adatoms on the noble metal surfaces have received considerable experimental and theoretical interest during the recent years (Torrente F. et al., 2007; Ziegler M. et al., 2008; Nanayakkara S. U. et al., 2007). The interactions have been explained by Friedel oscillations in the surface-state electrons, since they show the characteristic asymptotic dependence as  $E(r) = -A \sin(2k_F r + 2\delta) / r^2$  (Lau K. H. et al., 1978). Low-temperature STM studies have resolved the long-range interactions mediated by surface states at large interatomic separations (Repp J. et al., 2000; Knorr N. et al., 2002) and found local diluted hexagonal structures. In addition, the long-range interactions are also considered as the driving force for the formation of the superlattice (Silly F. et al., 2004; Negulyaev N. N. et al., 2006). Several theoretical studies have demonstrated that those indirect interactions can highly affect atomic motion and the growth processes despite the fact that they are relatively small. For example, Bogicevic *et al.* (Bogicevic A. et al., 2000; Ovesson S. et al., 2002) revealed a large increase of island density when the long-range interactions are taken into account. Fichthorn *et al.* (Fichthorn K. A. & Scheffler M., 2000) found that the long-range interactions between Ag adatoms on a strained Ag(111) surface are comparable to the diffusion barrier, which can significantly influence the surface diffusion and the growth morphology in a thin-film epitaxy. Combine with kinetic Monte Carlo simulations, Fichthorn *et al.* (Fichthorn K. A. et al., 2002, 2003; Merrick M. L. et al.,

2003) also found that the repulsive interactions make the island densities over an order of magnitude larger than those predicted by nucleation theory. In addition, the repulsive part of interactions can lead to monodisperse islands early in the deposition process.

The effect of the long-range interactions on fabricating nanostructures are divided into two branches: the formation of uniformly distributed islands such as quantum dots (Liu C. H. et al., 2006; Negulyaev N. N. et al., 2008) and the large scale of superlattice which consists of regularly monomers in hexagonal structures or nanochains (Silly F. et al.; Ding H. F. et al., 2007; Negulyaev N. N. et al., 2008; Negulyaev N. N. et al., 2009).

In the present work, the nucleation and island growth on (111) surface in considering the repulsive interactions are systematically investigated by kinetic Monte Carlo (kMC) simulations. The dependence of radius, the attenuation rate, the intensity of the repulsive interactions and the diffusion barrier on the island density is discussed in detail. Based on the study, a relatively clear understanding of the relationship between the island density and repulsive interactions is obtained.

## 2. The kMC simulation model

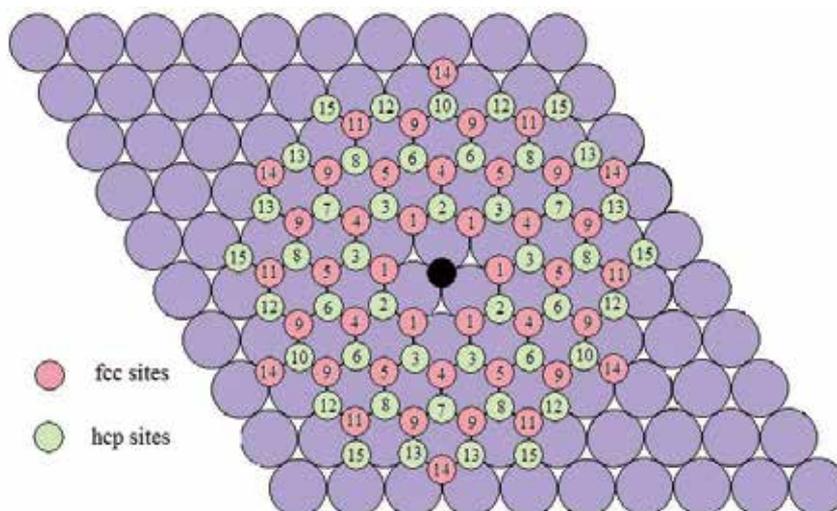


Fig. 1. Schematic illustration of the (111) substrate and possible adsorption sites. The big purple particles denote the (111) substrate. The black dot in the center denotes the position of one adatom on the substrate and other small dots indicate possible positions of a second adatom. The red dots represent fcc sites and the green dots represent the hcp sites. The number on the small dots indicates the  $i$ th neighbors from the central adatom.

As shown in Figure 1, the (111) substrate is represented as a triangular lattice with two non-equivalent adsorption sites (fcc and hcp) with a separation of  $a_0 / \sqrt{3}$  between the nearest sites, where  $a_0$  is the lattice constant of the substrate. Adatoms can only diffuse from the current fcc (hcp) site to one of the nearest empty hcp (fcc) sites at low temperature. The diffusion rate of an adatom from site  $k$  to site  $j$  on the (111) surface is calculated using the expression  $v_{k \rightarrow j} = v_0 \exp(-E_{k \rightarrow j} / k_B T)$ , where  $v_0 = 10^{12} \text{ s}^{-1}$  is the attempt frequency (Fichthorn K.A. et al., 2000),  $E_{k \rightarrow j}$  is the barrier between site  $k$  and site  $j$ ,  $k_B$  is the Boltzmann constant.

The influence of the adatom-adatom interactions on diffusion is included in the barrier which takes the form:  $E_{k \rightarrow j} = E_d + 0.5(E_j - E_k)$ . Here,  $E_d$  is the diffusion barrier for an isolated atom on a clean surface,  $E_{k(j)}$  is the sum of the pair interactions of the hopping adatom with all other adatoms in the system when the hopping adatom is at site  $k(j)$ . The difference of diffusion barriers between fcc and hcp sites and the effect of tri-adatoms interactions are neglected in our simulation. The simulation box is set to  $200 a_0 \times 200 a_0$  with the periodic boundary conditions. The deposition rate is fixed at  $F=0.01\text{ML/s}$  during the simulation. Therefore, the step interval between two continuous depositions is  $v_0 \exp(-E_d/k_B T)/(200 \times 200 F)$ . The simulation stops when the coverage reaches  $0.05\text{ML}$  and there is no isolated adatoms.

The specific configuration of the (111) surface in figure 1 shows that the central adatom is surrounded alternately by fcc (red dots) and hcp (green dots) sites. The repulsive barriers we defined here is either at hcp sites or fcc sites. Thus the shape of the repulsive rings is hexagonal. The numbers on each small site in Figure 1 mean the  $i$ th neighbors from the central adatom. The interactions between more than two adatoms are described by pairwise summation (Österlund L. et al., 1999; Fichthorn K. A. et al., 2003).

### 3. Simulation results and discussion

#### 3.1 The radius of the repulsive ring

Different species of adatom and substrate have different long-range interactions, including the intensity, the location of the repulsive barriers and the attenuation rate despite of the similar origin. The location of the repulsive barriers of Cu/Cu(111) is at  $12 \text{ \AA}$  (9th-11th neighbors) (Repp J. et al., 2000) while that of Ag on strained Ag(111) is at about 10th-13th neighbors (Fichthorn K. A. et al., 2000). Thus the effect of the location of the repulsive barriers to the island density is necessary to be studied. As shown in figure 1, around the central adatoms, the possible locations are alternant between fcc sites and hcp sites. In this study, the location of the repulsive barriers are assumed to be at the 2-3th, 4-5th, 6-7-8th, 9-11th, 10-12-13-15th, and 14-16-18th neighbors, respectively. In all these cases, there is only one repulsive barrier with unique intensity before the formation of island. Thus, the range of the repulsive barriers is the same and only effect of the distance from the adatom to the repulsive barriers (the radius of the repulsive ring) is considered.

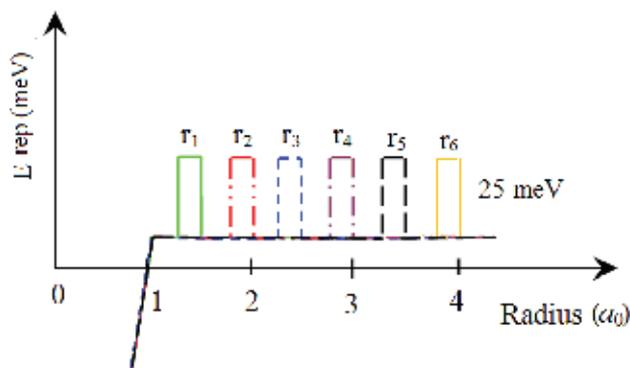


Fig. 2. Schematic illustration of the location of the repulsive barriers with unique intensity.  $a_0$  is the lattice constant of the substrate. The intensity of the repulsive barriers is set to  $25 \text{ meV}$ .

The schematic illustration of the locations of the repulsive barriers is shown in Figure 2. The intensity of the repulsive barrier is 25 meV, which is the same as the diffusion barrier. The substrate temperature is set to 20 k. In Figure 3(a), we show the relationship between the island density and the location of the repulsive barriers. With the increase of the radius of the repulsive ring, the island density increase quickly at first then reach the saturation when the barrier is at about 12th neighbors and at last decrease with the continuous increasing distance.

In the simulation, the adatom-adatom interactions are pairwise and the barriers around the island are the summation of the interactions comes from all the atoms in the island. Thus, the barrier around the island is larger than that around the isolated adatom, which makes the nucleation easier than the growth of the island. In addition, the barrier around the island depends on the radius of the repulsive ring even the size of the island is the same. An example is given in Figure 4. It is the potential energy map around a linear trimer with the repulsive barriers at 4-5th and 9-11th neighbors, respectively. We can see that the range of the repulsive barrier at the end of the trimer in Figure 4(b) is larger than that in Figure 4(a), which means that the monomers that overcome the repulsive barriers in Figure 4(b) will have more probabilities to fall back before forming a stable bond with the existing island. Beside the range of the repulsive barrier around the islands, adatoms with large radius of repulsive ring have higher probabilities to jump out to keep isolated when entered the repulsive ring of other adatoms.

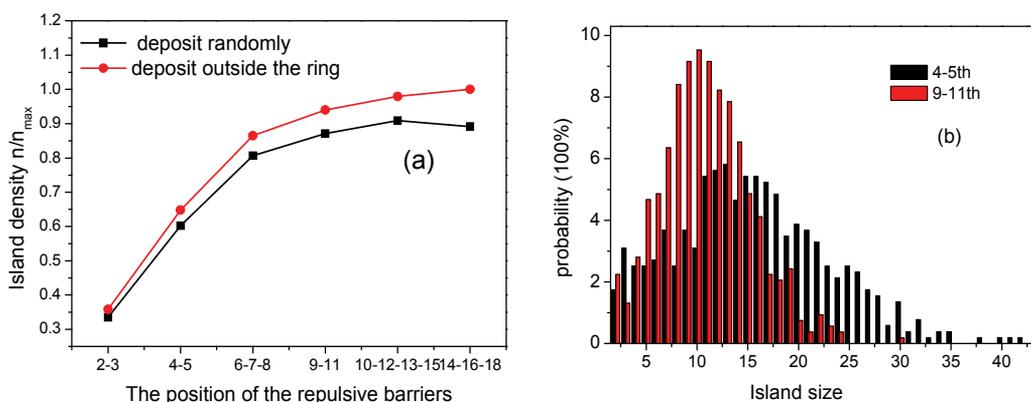


Fig. 3. (a) Island density as a function of the location of the repulsive barriers with randomly deposition and deposition only outside the repulsive ring. (b) Island size distribution with repulsive barriers at the 4-5th and 9-11th neighbors, respectively.

The decrease of the island density with the continuous increase of the radius comes from the deposition process. The probability to deposit inside the repulsive ring of other monomers or islands increases in square with the size of the repulsive ring. Spontaneously, the nucleation and growth occur easily when these events happen. To approve this, the atoms deposited inside the repulsive ring of other adatoms are cancelled until the distance between all of the adatoms and new adatom is larger than the repulsive ring. As shown the red curve in Figure 3(a), the island density not only increases monotonously until saturation but also has an up shift throughout the whole conditions studied.

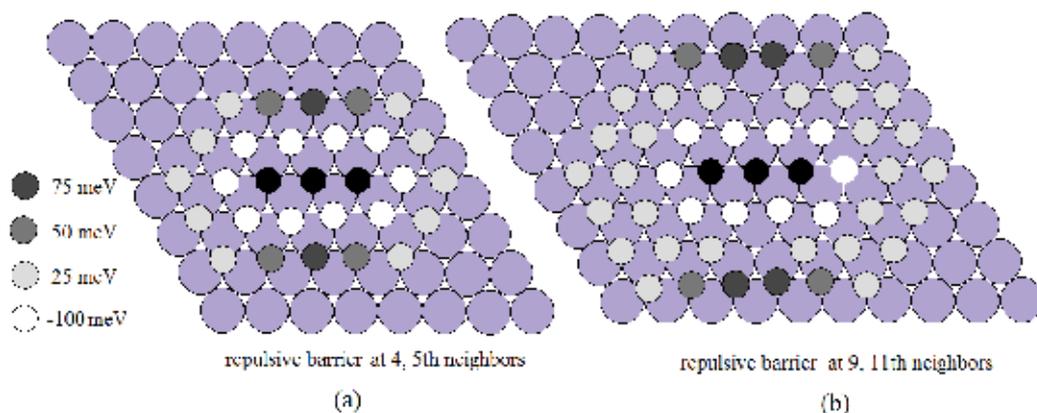


Fig. 4. Potential energy maps around a linear trimer with the location of the repulsive barriers at 4-5th neighbors and 9th-11th neighbors. The large purple particles represent the substrates, the three black dots in the center of the substrate represent the trimer. The gray dots in different colors mean different potential energy, the darker the larger of repulsive barriers.

Figure 3(b) is the results of the island size distribution with different radius of the repulsive ring. It shows that the distribution becomes sharper when the distance changes from 4-5th neighbors to about 9-11th neighbors. The average size of the island changes from 15 atoms to 11 atoms. This result implies that the position of the repulsive barriers is a candidate to modulate the size of the island to fabricate desired nanostructures.

### 3.2 The attenuation rate of the repulsive barriers

In the long-range interactions expression, another parameter varies in different materials is the attenuation rate of the repulsive ring. For example, the oscillation period on Cu(111) surface is about 15 Å (Repp J. et al., 2000) while on Ag(111) surface it is about 38 Å (Silly F. et al. 2004), which is more than twice longer than that of Cu. Thus the attenuation rate of the repulsive interactions is much small on the Ag(111) surface. In the previous case, we only investigated the radius of the repulsive ring with uniform attenuation rate (only one uniform barrier). In this part we will investigate the effect of the attenuation rate of the repulsive barriers on the island density. Due to the pairwise summation, it is too high for adatom to penetrate if high repulsive barrier is considered with small attenuation rate. So the largest repulsive barrier in this simulation is set to 4 meV at 4-5th neighbors.

The schematic illustration of the repulsive barriers with different attenuation rate is shown in Figure 5. The repulsive barriers decrease uniformly with the increase of radius by keeping the inner intensity fixed. Five different attenuation rates are studied and the results are shown in Figure 6. It can be seen that the island density is highly influenced by the attenuation rate of the repulsive barriers. The island numbers is more than 2 times larger when 5 attenuation rings are introduced instead of a single barrier with the same largest intensity. Based on the simulation model, the probability for one adatom to nucleate with another monomer is the same whether there is one large barrier or five small barriers if the largest barrier is fixed. The difference comes from the barriers around small islands. For example, in the five repulsive barriers condition, the potential energy at the end of the linear trimer in figure 4 is no longer uniform. It has grades and the largest barrier will have a climb

on the basis of uniform repulsive interaction, which makes the adatoms difficult to incorporate into the existing island. Therefore, the wavelength of the electronic gas on the metal surface is an important parameter affecting the growth of the island after nucleation.

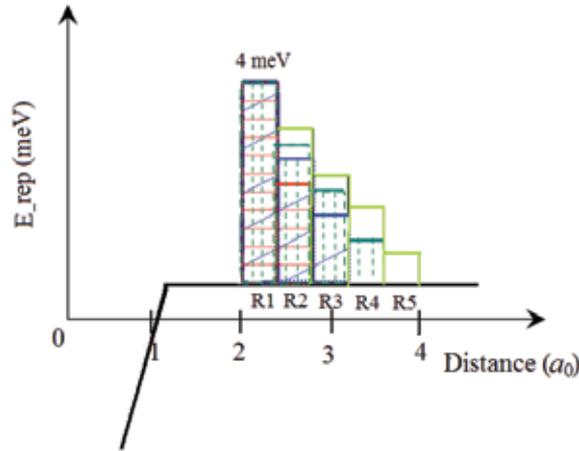


Fig. 5. Schematic illustration of repulsive barriers with different attenuation rate.

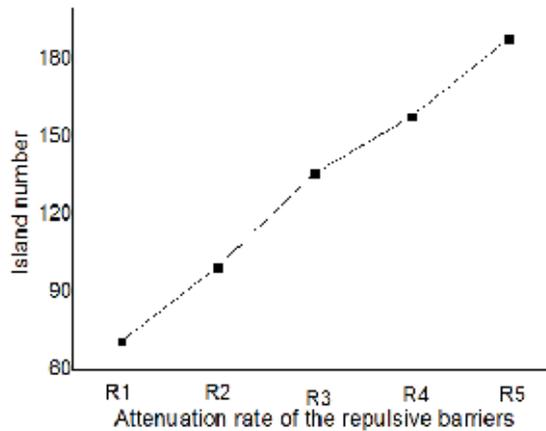


Fig. 6. The relation between the island density and the attenuation rate of the repulsive barriers.

### 3.3 The intensity of the repulsive rings

As Bogicevic *et al.* (Bogicevic A. *et al.*, 2000; Ovesson S. *et al.*, 2001) have already reported that there is a large increase of island density when the long-range interactions are taken into account. The increase of the island density originates from the repulsive part of long-range interactions around the adatoms which prevent the nucleation and the growth of the island. Here the specific effect of the intensity of the repulsive barrier on the island density is studied by fixing the diffusion barrier (25meV) and the location of the repulsive barrier (6-7-8th neighbors). The sketch map of the different repulsive barriers is given in figure 7. The intensities of repulsive barriers used are from 0 to 40 meV with an interval of 5 meV.

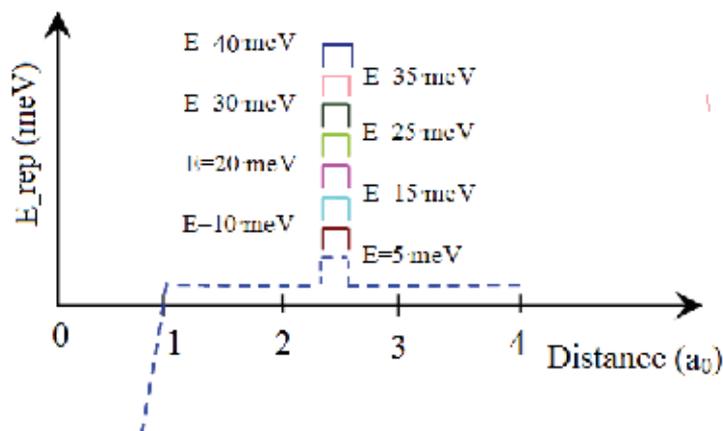


Fig. 7. Schematic illustration of the repulsive barriers around adatoms with the intensities from 0 to 40 meV. All of the locations of the barriers are at 6-7-8th neighbors.

According to the simulation model mentioned above, the influence of the adatom-adatom interaction on diffusion is included in the barrier taking the form:  $E_{k \rightarrow j} = E_d + 0.5(E_j - E_k)$ . So the probability to overcome the repulsive barrier of an monomer is as small as  $0.5 \exp(-E_{rep}/k_B T)$ . Figure 8 shows the simulation results of the island density with different repulsive barriers. We can see that the island density increase sharply at small repulsive barriers. The number of islands increases from about 52 to 100 when only a 5meV barrier is taken into account.

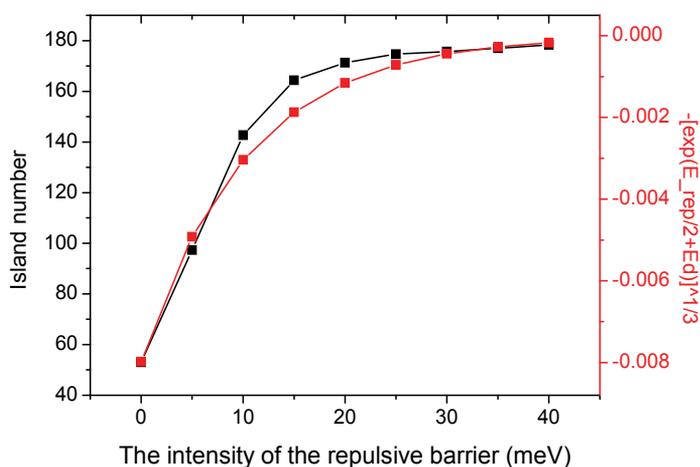


Fig. 8. The relation between the island density and the intensity of the repulsive barrier.

The increase rate of the island in figure 8 is compared with the expression  $-\left[\exp(-E_{rep}/2 - E_d)/k_B T\right]^{1/3}$ , which satisfies the mean-field theory if the  $E_{rep}/2 + E_d$  is regarded as the diffusion barrier. We find that the increase rate of the island density is nearly comparable to the exiting theory when the repulsive barrier is small. Thus the effect of the repulsive barrier can be roughly treated as an increase of the diffusion barrier when the repulsive barrier is small at low temperature and low coverage.

### 3.4 The diffusion barrier

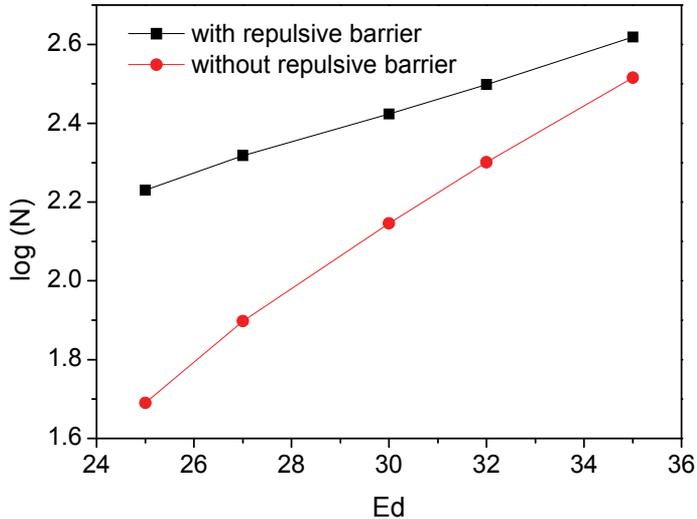


Fig. 9. Island density as a function of the diffusion barrier.

The diffusion barrier is another important parameter for the nucleation. The larger the diffusion barrier, the higher the temperature required for adatoms to diffuse and overcome the repulsive rings. In this part, the effect of the diffusion barrier on the nucleation is discussed after considering the repulsive barriers between adatoms. The results are given in figure 9. From the comparison whether considering the repulsive interactions or not, we find that both island densities are exponentially proportional the diffusion barrier. The difference is the slop of the line which decreases to one half when the repulsive barriers are considered. It means that the dependence of the island density on the diffusion barriers is no longer as sensitive as before. This result will be favored in heterogeneous growth where the diffusion barriers on the thin film are larger than that on the substrate.

## 4. Conclusions

Using kinetic Monte Carlo simulation, the effect of the repulsive part of the long-range interactions on the nucleation and island growth during thin film epitaxy is investigated. The parameters, including the radius, the attenuation rate, the intensity of the repulsive interactions, and the diffusion barrier of the substrate are discussed systematically. We find that the repulsive adatom-adatom interactions lead to higher island density than those without the repulsive interactions. The radius of the repulsive ring is an important factor determining not only the island density, but also the island size distribution. The slow attenuation repulsive rings can effectively block the growth of the existing islands. In addition, the island density depends sensitively on the intensity of repulsive barrier. Only a small repulsive interaction will highly increase the island density. Contrarily, the diffusion barrier is not as sensitive as that without repulsive interactions. The nucleation and growth mechanism which takes into account the repulsive interactions is propitious to layer-by-layer growth during the molecular beam epitaxy.

## 5. References

- Amar J. G. & Family F. 1995 Critical Cluster Size: Island Morphology and Size Distribution in Submonolayer Epitaxial Growth. *Physical Review Letters* 74, 2066-2069.
- Barth J.V., Brune H., Fischer B., Weckesser J. & Kern K. 2000. Dynamics of Surface Migration in the Weak Corrugation Regime. *Physical Review Letters* 84, 1732-1735.
- Bogicevic A., Ovesson S., Hyldgaard P., Lundqvist B. I., Brune H. & Jennison D. R. 2000. Nature, Strength, and Consequences of Indirect Adsorbate Interactions on Metals. *Physical Review Letters* 85, 1910-1913.
- Brune, H., 1998. Microscopic view of epitaxial metal growth: nucleation and aggregation. *Surf. Sci. Rep.* 31, 121.
- Brune H., Bales G.S., Jacobsen J., Boragno C. & Kern K. 1991. Measuring surface diffusion from nucleation island densities. *Physical Review B*, 60, 5991-6006.
- Brune H., Bromann K., Röder H., Kern K. 1995. Effect of strain on surface diffusion and nucleation. *Physical Review B*, 52, R14380-R14383.
- Ding H. F., Stepanyuk V. S., Ignatiev P. A., Negulyaev N. N., Niebergall L., Wasniowska M., Gao C. L., Bruno P. & Kirschner J. 2007. Self-organized long-period adatom strings on stepped metal surfaces: Scanning tunneling microscopy, ab initio calculations, and kinetic Monte Carlo simulations. *Physical Review B*, 76, 033409.
- Fichthorn K.A., Scheffler M., 2000. Island Nucleation in Thin-Film Epitaxy: A First-Principles Investigation. *Physical Review Letters* 84, 5371-5374.
- Fichthorn K. A., Merrick M. L. & Scheffler M., 2002. A kinetic Monte Carlo investigation of island nucleation and growth in thin-film epitaxy in the presence of substrate-mediated interactions. *Applied Physics A*, 75, 17-23.
- Fichthorn K. A., Merrick M. L., 2003. Nanostructures at surfaces from substrate-mediated interactions. *Physical Review B*, 68, 041404(R).
- Fischer B., Brune H., Barth J.V., Fricke A. & Kern K., 1999. Nucleation Kinetics on Inhomogeneous Substrates: Al/Au(111). *Physical Review Letters* 82, 1732-1735.
- Knorr N., Brune H., Epple M., Hirstein A., Schnieder M. A. & Kern K., 2002. Long-range adsorbate interactions mediated by a two-dimensional electron gas. *Physical Review B*, 65, 115420.
- Lau K. H. & Kohn W. 1978. Indirect long-range oscillatory interaction between adsorbed atoms. *Surface Science*, 75, 69-85.
- Liu C. H., Matsuda I., D'angelo M., Hasegawa S. J., Okabayashi J., Toyoda S. & Oshima M., 2006. Self-assembly of two-dimensional nanoclusters observed with STM: From surface molecules to surface superstructure. *Physical Review B*, 74, 235420.
- Merrick M. L., Luo W. W. & Fichthorn K. A., 2003. Substrate-mediated interactions on solid surfaces theory, experiment, and consequences for thin-film morphology. *Progress in Surface Science*, 72, 117-134.
- Nanayakkara S. U., Sykes E. C. H., Torres L. C. F., Blake M. M. & Weiss P. S., 2007. Long-Range Electronic Interactions at a High Temperature: Bromine Adatom Islands on Cu(111). *Physical Review Letters* 98, 206108.
- Negulyaev N. N., Stepanyuk V. S., Niebergall L., Hergert W., Fangohr H. & Bruno P., 2006. Self-organization of Ce adatoms on Ag(111): A kinetic Monte Carlo study. *Physical Review B*, 74, 035421.
- Negulyaev N. N., Stepanyuk V. S., Hergert W., Bruno P. & Kirschner J., 2008. Atomic-scale self-organization of Fe nanostripes on stepped Cu(111) surfaces: Molecular dynamics and kinetic Monte Carlo simulations. *Physical Review B*, 77, 085430.

- Negulyaev N. N., Stepanyuk V. S., Bruno P., Diekhöner L., Wahl P. & Kern K., 2008. Bilayer growth of nanoscale Co islands on Cu(111). *Phys. Rev. B* 77, 125437.
- Negulyaev N. N., Stepanyuk V. S., Niebergall L., Bruno P., Pivetta M., Ternes M., Patthey F. & Schneider W-D., 2009. Melting of Two-Dimensional Adatom Superlattices Stabilized by Long-Range Electronic Interactions. *Physical Review Letters* 102, 246102.
- Österlund L., Pedersen M. Ø., Stensgaard I., Lægsgaard E. & Besenbacher F. 1999. Quantitative Determination of Adsorbate-Adsorbate Interactions. *Physical Review Letters*, 83, 4812-4815.
- Ovesson S., Bogicevic A., Wahnström G. & Lundqvist B. I., 2001. Neglected adsorbate interactions behind diffusion prefactor anomalies on metals. *Physical Review B*, 64, 125423.
- Ratsch C., Zangwill A., Smilauer P. & Vvedensky D. D., 1994. Saturation and scaling of epitaxial island densities. *Physical Review Letters*, 72, 3194-3197.
- Repp J., Moresco F., Meyer G., Rieder K.-H., Hyldgaard P. & Persson M., 2000. Substrate Mediated Long-Range Oscillatory Interaction between Adatoms: Cu /Cu(111). *Physical Review Letters*, 85, 2981-2984.
- Silly F., Pivetta M., Ternes M., Patthey F., Pelz J. P. & Schneider W. D., 2004. Coverage-dependent self-organization: from individual adatoms to adatom superlattices. *New Journal of Physics*, 6, 16.
- Silly F., Pivetta M., Ternes M., Patthey F., Pelz J. P. & Schneider W. D., 2004. Creation of an Atomic Superlattice by Immersing Metallic Adatoms in a Two-Dimensional Electron Sea. *Physical Review Letters*, 92, 016101.
- Torrente F., Monturet S., Franke K. J., Fraxedas J., Lorente N. & Pascual J. I., 2007. Long-Range Repulsive Interaction between Molecules on a Metal Surface Induced by Charge Transfer. *Physical Review Letters*, 99, 176103.
- Venables J. A., 1973. Rate equation approaches to thin film nucleation kinetics. *Philosophical Magazine*, 27, 697-738.
- Venables J. A., Spiller G. D. T. & Hanbücken M., 1984. Nucleation and growth of thin films. *Reports on Progress Physics*, 47, 399.
- Ziegler M., Kröger J., Berndt R., Filinov A. & Bonitz M., 2008. Scanning tunneling microscopy and kinetic Monte Carlo investigation of cesium superlattices on Ag(111). *Physical Review B*, 78, 245427.

# Monte Carlo Methodology for Grand Canonical Simulations of Vacancies at Crystalline Defects

Dôme Tanguy

*CNRS, UMR 5146, Ecole des Mines de Saint-Etienne  
France*

## 1. Introduction

The design of new materials and the optimization of the existing ones require more and more knowledge of the elementary processes underlying the macroscopic properties. Computer simulations have become, together with ever finer experimental technics, the modern tools for probing these mechanisms. This paper focuses on the development of Monte Carlo simulations, at the atomic scale, of vacancies in crystals. These defects have been extensively studied, in their isolated state, because they are the vectors of diffusion in solids. Their concentration and dynamics determine the kinetics of most phase transformations and thermal annealings which enter the processes of the production of materials, for example metallic alloys, or surface deposits for microelectronic applications. They can also contribute to the loss of mechanical properties. For example, in irradiated steels, the clustering of vacancies induce the formation of loops which harden the matrix and, at the same time, their diffusion to the grain boundaries lead to all sorts of segregations that sometimes reduce their cohesion. The combined effect of a hard matrix and weak interfaces can lead to the premature formation of cracks. It is therefore not only important to model vacancies in a hole range of temperatures and concentrations in perfect crystals, but also at pre-existing defects, like grain boundaries and dislocations.

The methodology presented is inherited from statistical mechanics. Molecular Dynamics (MD) (Allen & Tildesley (1991)) is the method of choice if the details of the trajectories of the particles in the system are needed. The amount of physical time that can be simulated (of the order of the nano second) is often too limited to give enough statistics to measure the property of interest. Kinetic Monte Carlo (Landau & Binder (2000); Soisson et al. (1996); Dai et al. (2005)) is event based. It eliminates all the details of the trajectory and keeps only the jumps from one local minimum of the energy to another one. In its simple form, a limited list of the most important events is provided at the beginning of the simulation, together with the list of rates and the particles are constrained to be on a rigid lattice. It can be refined to build the list on the fly (Henkelmann & Jónsson (2001)) or to get the events from MD (Sørensen & Voter (2000)). A last class of methods is the one where MD is accelerated (Voter et al. (2001)), for example, by the use of a bias in the interactions (Wang et al. (2001)) which does not modify the saddles in between the local energy minima, but reduces the waiting time in the basins. Each method has its limitations: KMC, on a rigid lattice, can not treat realistic diffusion mechanisms with collective movements and relaxations (for example, if the system has different components

with marked different sizes), Accelerated Dynamics is sometimes still limited to short time scales when many low barriers are present.

We feel the need to develop some tools which lie in between rigid lattice methods and MD by borrowing some of their characteristics. First of all, this study is focused on simulating equilibrium configurations, but in rather complex geometries like grain boundaries. Monte Carlo and Molecular Dynamics are used, together, to treat interstitial-vacancy clusters and vacancies at grain boundaries. This effort can only be considered as a first step since, most of the time, out of equilibrium situations are met in experiments. It means that the work should be oriented towards rate calculations in the future.

The paper is organized as follows: A review of equilibrium Monte Carlo simulations in the Grand Canonical ensemble is given. Insertion/deletions moves are presented. They are used to obtain a fluctuating number of particles in the system. The biases that have been developed to extend this method to dense systems are also presented. Next, we detail our model which is an intermediate between a rigid lattice and a continuum model. An application to the simulation of thermal vacancies at a grain boundary is given. Then the model is enriched to treat also interstitial solutes. Vacancy-hydrogen clusters are simulated in a perfect crystal. The focus is on the design of cluster moves. The method is extended to grain boundaries. H segregation is shown as an example. The problem of the slow convergence is discussed in the case of vacancy-hydrogen co-segregation. Finally, an algorithm is developed to solve this problem and is applied to the ordering of vacancies alone in a grain boundary.

## 2. Simulations in the Grand Canonical ensemble, in dense systems

The method for simulating the Grand Canonical ensemble is briefly reviewed, first in the context of the low density systems such as gases or low density fluids. Next, the modifications that were made to extend this method to dense liquids or the hexatic phase are presented.

Consider a set of  $N$  labeled classical particles, in a volume  $V$  at temperature  $T$ . A microstate of the system is characterized by the continuous positions and momenta of the particles:  $\{q_i\}^N \{p_i\}^N$ , where  $q$  and  $p$  are vectors and  $i$  is the label of the particle. One microstate is associated to an infinitesimal volume of the space  $(\{q_i\}^N, \{p_i\}^N)$ , proportional to  $(\{dq_i\}^N, \{dp_i\}^N)$ , the uncertainty of the measure of the positions and impulsions. Consider that this system is in equilibrium with a reservoir with which it can exchange particles, such that the chemical potential is fixed (all particles are considered to be of the same chemical nature). The probability that the system be in the microstate  $(N, \{q_i\}^N, \{p_i\}^N)$  is:

$$p(N, \{q_i\}^N, \{p_i\}^N) \{dq_i\}^N \{dp_i\}^N = \frac{1}{\mathcal{Q}(\mu, V, T)} \frac{1}{N!} \exp(-\beta(H(\{q_i\}^N) + \frac{1}{2} \sum_{i=1}^N 1/2m_i p_i^2 - N\mu)) \frac{\{dq_i\}^N \{dp_i\}^N}{h^{3N}} \quad (1)$$

where  $\mathcal{Q}(\mu, V, T)$  is the partition function:

$$\mathcal{Q}(\mu, V, T) = \sum_{N=0}^{+\infty} \int \frac{1}{h^{3N} N!} \exp(-\beta(H(\{q_i\}^N) + \frac{1}{2} \sum_{i=1}^N \frac{1}{2m_i} p_i^2 - N\mu)) \{dq_i\}^N \{dp_i\}^N \quad (2)$$

In the case where the quantities of interest do not explicitly depend on the velocities of the particles, only the configurational part of the density can be considered and the kinetic part is integrated by hand. Equation 1 becomes:

$$p(N, \{q_i\}^N) \{dq_i\}^N = \frac{1}{Q(\mu, V, T)} \frac{1}{N! \Lambda^{3N}} \exp(-\beta(H(\{q_i\}^N) - N\mu)) \{dq_i\}^N \quad (3)$$

where  $\Lambda = \sqrt{(h^2/2\pi mkT)}$ .

In order to let the number of particles fluctuate (since the extensive variable V is fixed and that the particles cannot overlap strongly, the number of particles oscillates around an average value), Monte Carlo simulations can be performed with a specific trial move: the insertion/deletion move.

The probability of the transition from a state o (old) to a state n (new) is decomposed:  $p_{o \rightarrow n} = \rho_o \alpha_{o \rightarrow n} acc_{o \rightarrow n}$ , with  $\rho_o$  the probability that the system, in equilibrium, is in state o (Eq. 3);  $\alpha$  is the probability to propose the transition (trial) and  $acc$  is the probability to accept the transition. Let's consider that the n state is obtained by inserting, at random, a particle in volume V. The probability to pick the new position  $q_{N+1}$  within the box of length  $dq_{N+1}$  is  $dq_{N+1}/V$ , if we imagine that physical space is decomposed in small boxes of length  $dq$ . The position of all the other particles is left unchanged. Once the new particle is inserted, the particles need to be relabeled. We take one label at random between 1 and (N+1) for the new particle and we proceed like that for all particles sequentially. The new labeling is obtained with probability  $1/(N+1)!$ . The reverse move is obtained by selecting the former "new" particle among the others. This is done with probability  $1/(N+1)$ . The particle is removed and the configuration is relabeled. This gives N! possibilities. The probability to come back to the old configuration, with the same labeling is  $1/N!$ . A sufficient condition to ensure that the Markov chain constructed by proposing insertion/deletions produces the desired equilibrium distribution of states (coherent with  $\rho$  eq. 3) is "detailed balance":

$$\rho_o \alpha_{o \rightarrow n} acc_{o \rightarrow n} = \rho_n \alpha_{n \rightarrow o} acc_{n \rightarrow o} \quad (4)$$

where  $\alpha_{o \rightarrow n} = dq_{N+1}/V \times 1/(N+1)!$  and  $\alpha_{n \rightarrow o} = 1/(N+1) \times 1/N!$ . Substituting Eq. 3 and the values of  $\alpha$  discussed in the text gives:

$$\frac{1}{\Lambda^{3N} N! Q} \exp(-\beta(H(\{q_i\}^N) - N\mu)) dq_i^N \frac{dq_{N+1}}{V (N+1)!} acc_{o \rightarrow n} = \frac{1}{\Lambda^{3(N+1)} (N+1)! Q} \exp(-\beta(H(\{q_i\}^N, q_{N+1}) - (N+1)\mu)) dq_i^N dq_{N+1} \frac{1}{(N+1) N!} acc_{n \rightarrow o} \quad (5)$$

$$\frac{acc_{o \rightarrow n}}{acc_{n \rightarrow o}} = \frac{V}{(N+1)\Lambda^3} \exp(-\beta(H(\{q_i\}^N, q_{N+1}) - H(\{q_i\}^N) - \mu)) \quad (6)$$

The new configuration is accepted according to the Metropolis rule (Allen & Tildesley (1991)). This algorithm has been used to simulate the equilibrium between a gas and a liquid (Adams (1975); Rowley et al. (1976)), generalized to Coulombic systems (Valleau & Cohen (1980)) and also applied to electrical double layers (van Meegen & Snook (1980)).

The critical part is the random insertion. When the density becomes large, the probability that the trial position gives a large overlap with an existing particle increases. The energy variation becomes large and the acceptance ratio drops. The sampling becomes inefficient. Mezei (Mezei (1980)) proposed to bias the insertions by detecting cavities in the system and extended the simulations to dense liquids. These ideas were revisited and extended by Swope (Swope &

Andersen (1992; 1996)). His algorithm is briefly presented. An interesting use of a grid is made to design an efficient insertion/deletion move, in the spirit of Mezei.

The idea is to attempt an insertion only at a position where it has a high probability to be accepted, respectively attempt a deletion of a particle that has a non negligible chance to be successful. This choice depends on the "old" configuration and also on the "new" configuration, because of the detailed balance condition. In order to construct the move, the volume of the system is decomposed in identical small cubic cells (the size is a fraction of the nearest neighbor distance). Their center is also considered. This decomposition is made at the beginning of the simulation and is not modified. The most relevant cells are extracted and called ID cells (ID: insertion deletion). The metropolis criterion contains the number of filled ID cells and the number of empty ID cells in addition to the usual energy variation term. There is a lot of flexibility in the definition of an ID cell. Swope chose, for the example he treated, to proceed in two steps:

- A geometric criterion: an ID cell, independently of whether it is occupied or not, has no particles closer than a fixed radius to its center. This is to avoid strong overlaps and to limit the number of ID cells in a computationally cheap way.
- An energy criterion: in the cells that satisfy the geometric criterion, a particle is inserted at their center (a fixed position, to satisfy reversibility) and the energy variation  $\Delta H$  is calculated (if a cell is occupied, the contribution of the particle is not considered). The cell is considered an ID cell if  $\Delta H$  is in a narrow range, which is common to the insertion and the deletion.

This energy range is chosen to increase the acceptance rate. Once the ID cells are defined and selected for the move, the insertion is performed at random in the ID cell if it is empty. Otherwise, the particle is removed. For the fluid that was studied, modeled by a twelfth power repulsive interaction, 20% of acceptance could be reached, which is remarkably high.

This algorithm uses a lattice (the cell centers), even if the system is not necessarily crystalline, which is original. The energy criterion, on the other hand, is quite expensive since it requires evaluation of many variations of energies.

### 3. Dealing with vacancies

Defining a vacancy implies that some degree of order, even only local or metastable, exists. A lattice structure can be defined by averaging the positions of the particles over a time scale smaller than the typical time scale for the disappearing of this local order. A vacancy is then an unoccupied lattice site. If, in simple crystalline structures, like fcc or bcc, the vacancies are usually well localized, it is not necessarily the case when crystalline defects are present. A spectacular example is given in (Denkowicz et al. (2008)). Atomic scale simulations of Cu-Nb multilayer systems, with layers being 4nm thick, were conducted. The Cu-Nb interfaces have several dense arrays of misfit dislocations. When introduced, vacancies delocalize and trigger complex dislocation rearrangements. As a consequence, the interface energy can be reduced by incorporating up to 5% of vacancies (the reference structure is the one obtained by sticking perfect layers, with different misorientations). This is a particular example where, by construction, the initial structure is stressed. In the case of a grain boundary, the system is trapped in a metastable state by the application of macroscopic constraints: the misorientation and the relative translation of the constitutive crystals. When looking for the ground state, atoms are removed until the lowest excess energy is found. It is common that several arrangements have close energies so that several structures coexist at finite temperature.

$\Sigma 5(210)[001]$  symmetrical tilt boundary has this property and can oscillate between two structures by migration of the interface (shear and translation perpendicular to the GB plane) or by absorption/emission of a vacancy per structural unit. It is therefore not straightforward to define a lattice for the GB core structure that can be suited to define vacancies. Furthermore, the localized state is not the only one for the vacancy: delocalization-re localizations are frequent during diffusion events. They can be simple exchanges with first neighbors or complex collective moves (Sørensen et al. (2000)), see (Suzuki & Mishin (2005)) for a review. Furthermore, when simulating crystalline systems, the initial size and shape of the box, which typically contains an integer number of elementary cells, imposes a strong constraint on the number of lattice sites which is not compatible with the equilibrium concentration of vacancies. These constraints, discussed by Swope (Swope & Andersen (1992)), can affect drastically the results of simulations of transitions between phases of very different structures (for example: melting (Bagchi et al. (1996))). In our case, large concentrations of vacancies, with a large number of lattice sites are involved (for a reason explained later). The impact of the constraint on the average concentrations of vacancies are not dramatic. Nevertheless, the issue of the unphysical distortions, also discussed in Swope's papers, are relevant and we do check the structures for them.

We use a model which is intermediate between a rigid lattice (Ising) and continuum. The particles are no longer represented by their positions respective to an origin but by a displacement relative to a lattice node (Tanguy & Mareschal (2005)). Node occupations (analogous to the spins of the Ising model) are defined:  $p_i = 1$  if node number  $i$  is occupied by a particle,  $p_i = 0$  if it is empty (vacancy). Furthermore, the displacements are confined to the Voronoi cells around the nodes. This constraint is natural in solids. No self interstitials are allowed. This is also acceptable if the crystal is stable and in equilibrium. A similar model was used for simulating the phase diagram of bulk Si-Ge, using a diamond lattice (Dünweg & Landau (1993)). Although the authors present their model as being able to handle vacancies, the insertion/deletion is not presented and the calculations were done, in practice in the semi-grand canonical ensemble.

With these considerations, a microstate is defined by the set of the occupancies and the displacements, when the nodes are filled:  $(\{p_i\}^M, \{\vec{u}_i\}^N)$ . There is one extensive variable in addition to the traditional ones which is the number of nodes  $M$  ( $N$  is the number of particles).  $M$  can be used to free the volume as discussed in (Tanguy & Mareschal (2005)). The partition function is:

$$\mathcal{Q}_c(M, \mu, V, T) = \sum_{N=0}^M \sum_{\{p_n\}} \frac{1}{\Lambda^{3N}} \int_{W.S} d\vec{u}^N \times \exp(-\beta(\mathcal{H}(\{p_n\}, (\vec{u})^N) - N\mu)) \quad (7)$$

where  $\mu$  is the chemical potential. The second sum (over the set of occupancy numbers) represents all the possible arrangements of the vacancies on the nodes of the lattice. Note that the integration of the displacement is over the volume of the Wigner-Seitz cell  $W.S$ . Monte Carlo simulations are done according to the density defined by Eq. 7 by:

- Proposing displacements to the particles, within their cell. If a random displacement increment brings the particle out of its cell, the new cell is identified. If it is empty, the occupancies are exchanged. If it is occupied, the move is considered as an attempt to have a double occupancy, which is not possible. A warning is issued and the move is abandoned.

In practice, it never happens since  $T$  is not close to the melting point. Nevertheless, the possibility is discussed below.

- The number of particles fluctuates. Insertion/deletions are proposed. First a node is selected at random. If it is occupied, the particle is deleted (the occupancy is set to 0). If it is not occupied, a random position is selected in the cell and the particle is inserted. Detailed balance constrains the acceptance in a similar way to Eq. 6:

$$\frac{acc_{A \rightarrow B}}{acc_{B \rightarrow A}} = \exp(-\beta(\mathcal{H}_B - \mathcal{H}_A + \mu)) \frac{\Lambda^3}{W.S} \quad (8)$$

where  $A$  is the initial state with  $N$  particles and  $B$  is the trial state with  $N-1$  particles.

- Exchanges are also proposed between a vacancy and a particle. If the cells have exactly the same shape (which is the case in the perfect lattice, but not in the grain boundary), the displacement of the particle is preserved. The exchanges are necessary to speed up the ordering of the vacancies.

This model is interesting because:

- It provides a simple and unambiguous definition of vacancies.
- Relaxations due to the presence of crystalline defects are taken into account (the displacement moves lead to the relaxation of the defects).
- The cell decomposition of space is used as insertion/deletion cells in the spirit of Swope.
- The lattice can be used to design cluster moves that speed up convergence.
- Lattice models, in the mean field approximation, can be used to check convergence in the low vacancy concentration limit.

The model is used in (Vamvakopoulos & Tanguy (2009)) to simulate thermal vacancies, at high temperatures, in the  $\Sigma 33(554)[110]$  symmetrical tilt boundary (Fig. 1). Because of the displacements, the equilibrium vacancy concentrations take into account the vibrational entropy which largely influences the formation energies in the different sites of the grain boundary. The Monte Carlo results are compared to free energy calculations using the Widom insertion method, which enabled to check the convergence. Clusters of vacancies, up to 5 vacancies along the tilt axis, are observed. Nevertheless the convergence is difficult because of the strong relaxations around the vacancies in the grain boundary. A solution for this problem is presented in the section concerning "Hybrid Monte Carlo".

It is natural to question the motivation for ignoring the self-interstitials in the case of the grain boundary. First, the systematic study of grain boundary diffusion (Suzuki & Mishin (2005)), show that in high energy boundaries, vacancies and self interstitials have similar formation energies and therefore, in equilibrium, similar concentrations. Furthermore, they are both involved in diffusion at low temperature. We have done the choice to focus on vacancies because we plan to work with large concentrations of vacancies, at low temperature. In this case, the influence of the low concentration of equilibrium self-interstitials is neglected. Note that large distortions are supposed to be handled by the lattice as it is, i.e. by the combination of a vacancy and the distortions of the neighbors within the limit of their cells. If it is not enough (attempts for double occupancies are tracked by the code), interstitial sites should be included. It would induce a refinement of the Voronoi decomposition because these new sites would be added to the regular lattice sites. Because space is re decomposed, the addition of extra sites, with extra occupancies, do not induce redundancies in the way the microstates

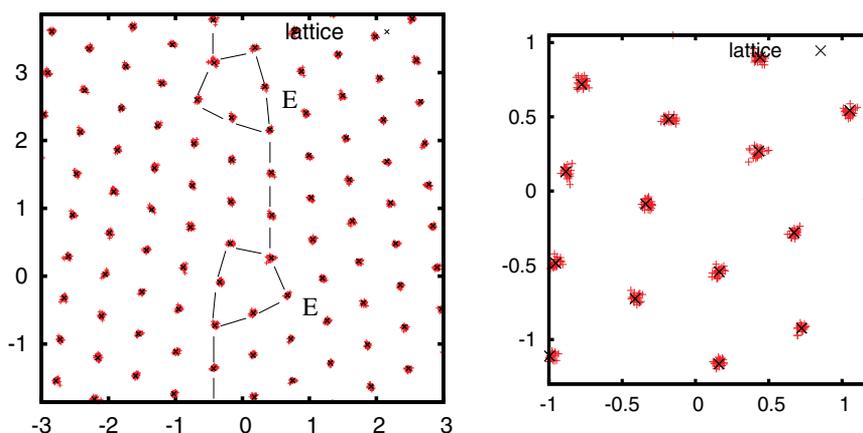


Fig. 1. Projection of the structural unit of the  $\Sigma 33(554)[110]$  symmetrical tilt boundary. (x) represents the position of the nodes. (+) are the actual positions of the atoms at thermal equilibrium, without vacancies, at  $T=200\text{K}$ .

are represented. The new decomposition might not be relevant and, if the original lattice is no longer adapted, it would be easier and more meaningful to use a refined cubic grid such as the one used by Swope, at least in the GB core region. If this is necessary, the concept of vacancy should be abandoned.

#### 4. Interstitial-vacancy clusters simulations in the Grand Canonical ensemble

There are different conditions where large concentrations of vacancies can be found. These can be heavily out of equilibrium: irradiation, intense localized plastic deformation (fatigue), oxidation (in particular alloys where one component is preferentially oxidized, the depleted zone is a direct evidence of enhanced diffusion, probably due to large vacancy concentrations). There exist also specific conditions where this can be observed, at equilibrium, in the case of a large binding between the vacancy and the solute introduced in the crystal. It is the case of many metal-hydrogen systems (Fukai (1993); Fukai & Ôkuma (1994)). In Ni, under high partial pressure of hydrogen and high temperature, the vacancy concentration can be as high as 25%. The origin of the enhanced equilibrium vacancy concentration lies on a significant binding and on the possibility of multiple occupancy of the vacancy by the solute. Indeed, it is known experimentally for a long time (Bugeat et al. (1976)) that H is not centered in the vacancy. In the eighties, Effective Medium Theory came to a good agreement with experiments and gave a simple and clear picture of H in metals (for a review see (Nørskov & Besenbacher (1987))): the minimum potential energy for H in the vacancy is intermediate between the geometric interstitial position and the center of the vacancy. Multiple occupancy is possible, but the H-H interaction in the vacancy is repulsive. This picture has been confirmed since then by *ab initio* calculations in many metals: Fe (Tateyama & Ohno (2003)), Al (Gunaydin et al. (2008); Wolverton et al. (2004)), W (Liu et al. (2009)), Be (Ganchenkova et al. (2009))... At zero temperature, multiple occupancy can lead to negative formation energies for the  $VH_n$  (a vacancy and  $n$  trapped hydrogen). Nevertheless, at finite temperature, the probability that H is actually in the vacancy depends on the concentration of hydrogen remaining on the regular interstitial sites (the chemical potential of H). It is therefore not straightforward

to conclude, from zero K binding energies, that  $VH_n$  clusters are stable. There can be some misunderstanding with regards to the experimental conditions: if the vacancy is forced into the crystal (by irradiation for example, or by corrosion) and if H is also present, it will segregate to the vacancy because of its much higher mobility. It might happen that the cluster is much less mobile than the free vacancy and that it gives the impression that the vacancies are stabilized by H (because they don't annihilate within the time scale of the experiment). It does not mean that the clusters are thermodynamically stable.

In the following, the equilibrium Monte Carlo method is extended to simulate the existence domain ( $C_{H,T}$ ) of vacancy clusters in the conditions where the chemical potential of the metal is imposed (i.e. the crystal is free to eliminate or create vacancies, as if the annealing was lasting for an infinite time in the experiment). The reason why the focus is on equilibrium simulations, and not on kinetic simulations (that would certainly be more relevant to the experimental conditions) is because sampling properly the equilibrium states and getting enough statistics, is already far from being granted, while equilibrium MC gives the advantage to use unphysical moves to get better convergence. This is the theme that is treated now.

#### 4.1 Grand canonical simulations in a perfect crystal

In a similar way to what is done for the metal, a lattice is introduced to handle interstitial solutes (Tanguy & Mareschal (2005)). In the fcc structure, it is composed of one octahedral site and two tetrahedral sites per fcc site. Space is decomposed in Voronoï cells based on this lattice. The interstitials are referenced by an occupation number and a displacement from the node of the cell where they belong. Phase space is extended and sampled by Monte Carlo moves: displacements and exchanges between occupied and empty nodes.

Let's imagine that a vacancy is formed. The exchange moves bring hydrogen in it and an equilibrium is created between the vacancies, the H which multiply occupy them and the remaining H on bulk interstitial sites. Suppose an exchange move is attempted for a vacancy containing  $n$  H. After the particle is brought on the empty site, the trial state will have one more free vacancy and  $n$  more isolated H on bulk sites. The energy variation would be of the order of  $n \times E_b$  where  $E_b$  is the binding energy. For H in Al,  $E_b \approx 0.3eV$ , which, if  $n = 3$ , gives  $\Delta E = 0.9eV$  and the probability to accept the move (metropolis) is  $exp(-0.9/kT) \approx 10^{-15}$  at  $T=300K$ . The direct exchange is never accepted. Not only the ordering of the vacancies is not described, but also the average occupancy (average  $n$  per vacancy) is not correct because vacancy clusters  $V_2...V_m$  do not trap H in the same way as an isolated vacancy.

A solution is to perform "cluster moves" i.e. to exchange a vacancy and the H atoms it contains, at the same time. Before we do this, we have to come back to the way the microstates are defined. For symmetry reasons, the tetrahedral sites coincide with the corners of the Voronoï polyhedron (and the octahedral site sits on an edge). It means that two vacancies in first neighbor position share 2 tetrahedral sites. If they were occupied, the H would relax towards the center of the vacancy and therefore, a T site can be occupied by 2 H atoms if each of them relaxes towards a different vacancy. It is possible that not allowing multiple occupancy prevents physical states to appear in the simulation. For this reason we decided to "split" all the interstitial sites. To each fcc site is associated 8 T and 6 O sites. Each of them is degenerate: the same T site appears 4 times but associated to four different atoms (the corners of the tetrahedron). To avoid redundant referencing of the microstates (which could be tolerated if corrected in the partition function), we also split the interstitial Voronoï cells. For example, consider a T site associated to a vacancy. It will be occupied only if the H is in the intersection of the Voronoï cell of the vacancy and of the T site. For the H case, it is

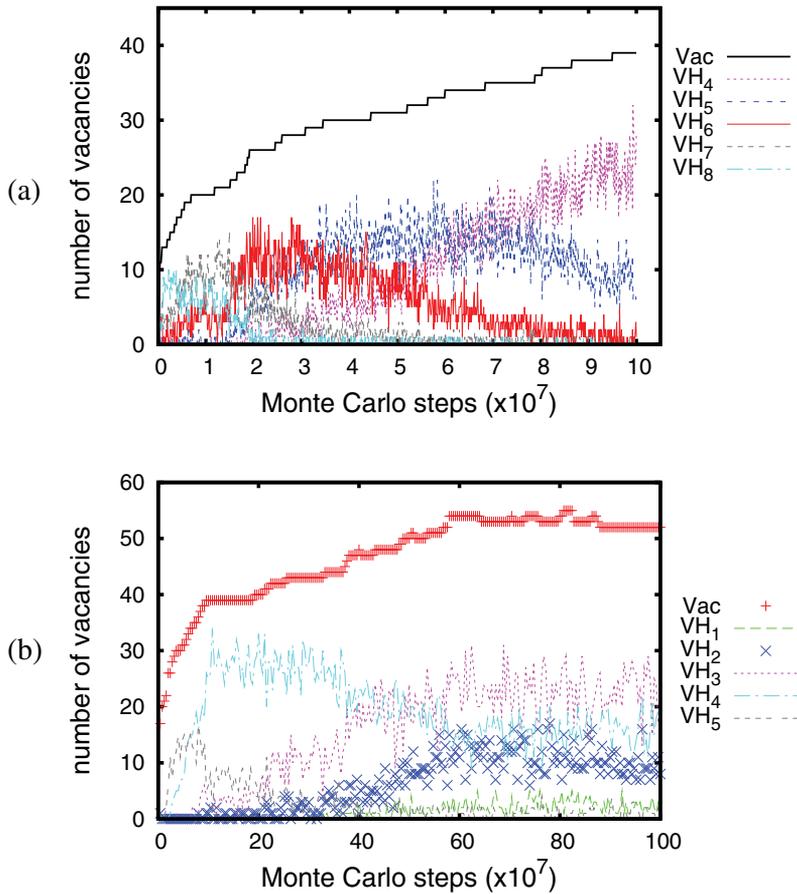


Fig. 2. Slow convergence of the total vacancy concentration and of the distribution of  $VH_n$  clusters in a perfect Al crystal containing 1%H at  $T=400\text{K}$ . The calculation involves: displacements of Al and H, insertion/deletions, volume changes, exchanges of H and cluster moves.

an efficient decomposition of space, because the H is physically relaxed towards the center of the vacancy. Defining a cluster is straightforward: it is the vacancy and the H belonging to the nodes it contains. Since, in the bulk, all the fcc sites are equivalent, the cluster move is reversible.

The Al-H system is used as a test system (Tanguy & Mareschal (2005)) because of the low formation energy of the vacancies (0.7eV) and the large segregation energy (-0.3eV). To test the cluster move and obtain a large concentration of vacancies, the total (including trapped H) concentration of hydrogen is 1%. The reader is warned that the calculations were done with a potential which overestimated the vacancy - hydrogen binding: -0.55eV instead of -0.3eV as is given nowadays by ab initio calculations. The results are therefore only illustrative of the method (a new potential, the one used for intergranular segregation, shows that more than 1%H is necessary to obtain large concentrations of vacancies at 400K). Figure 2 illustrates the slow convergence of the simulation. At the beginning of the simulation (Fig. 2a), the

concentration of vacancies increases. Because of the H exchange moves, the newly created vacancies are highly multiply occupied (up to 8 H per vacancies), but as more and more vacancies are created, the average occupancy decreases. Nevertheless, it is only when the vacancy concentration stabilizes and that the ordering of the vacancies is established that the true distribution of H is obtained (Fig. 2b:  $VH_2$ ,  $VH_3$  and  $VH_4$  are the dominant species). The presence of  $VH_2$  is really the signature of the ordering because it is when two vacancies are in first neighbor position that they can share 2 H (Tanguy & Mareschal (2005)).

This example highlights the importance of a good description of the ordering of the vacancies if the proper trapped hydrogen concentrations are to be measured.

#### 4.2 Hydrogen segregation to a grain boundary

Are hydrogen-vacancy clusters really involved in the hydrogen damage of metals? If this is the case, the first place to look for these objects is where they can be formed in large quantities: at crack tips, in the dislocation core or at grain boundaries (GBs). Indeed, these defects are preferential sites for vacancy formation (low formation energy) and for H segregation. Let us first focus on the segregation of H alone. Lattice sites are defined in the core of the GB (Fig. 1). Keeping in mind that cluster moves are necessary when vacancies are present, we decided to keep the same number of interstitial sites (8T and 6O) per metal site as in the bulk. The interstitial sites are initially created at the same relative distance from the metal than in the bulk i.e. permutations of  $(+/-0.25,+/-0.25,0)$  for the T site and  $(+/-0.5,0,0)$  for the O site. But, because the fcc structure is disrupted by the presence of the GB, these sites are no longer sit on the borders of the Voronoï cells of the metal. This geometric construction leads a large number of different sites, many of them quite close from each other. The decomposition in Voronoï cells of the interstitial network becomes unnecessarily complex. To solve this problem, all the geometric sites that are closer than a radius  $r_{int} = 0.215a_0$  are merged together on one site (the new position of the site is the average over the whole cloud of initial interstitial sites). After this procedure is done, a simple lattice, suitable for the definition of clusters, is obtained and used for H exchange moves and confined H displacements, just like in the bulk.

The reader who is not interested in details can skip this paragraph and the next one. The problem is that the interstitial sites are no longer equivalent, i.e. that their Voronoï cells don't have the same shape. So it is not always possible to perform the exchange and keep the same displacement. There are, at least, two solutions. First solution: the displacements are always taken at random in the new cell, which means that the ratio of the volume of the old and the new cell has to be taken into account in the metropolis criterion (this is the implementation that is used in Fig. 3). Second solution: it is easy to define the biggest sphere contained in all the Voronoï cells (just take half the smallest distance between two interstitial sites). Then, if an H particle is to be exchanged with a vacancy and its initial displacement is within this sphere (it is very often the case), then the displacement is conserved. Otherwise, a displacement is taken at random within the new Voronoï cell, but outside the sphere. Again, the ratio of the volumes (Voronoi volume - the sphere volume) must be included in the metropolis criterion. Figure 3a and b show a typical configuration obtained after equilibration with only displacement moves (H and Al) and exchange moves for H. The calculation is done at fixed volume and fixed total number of particles.

The next concern is about multiplicity of labelling of microstates. In the bulk, the interstitial sites are split between the different fcc Voronoï cells to have more possibilities for the H in the vacancies. This argument is still valid in the case of the intergranular sites. But this time, the interstitial Voronoï cells are not split, i.e. the same cell can contain several H. Of course, due to

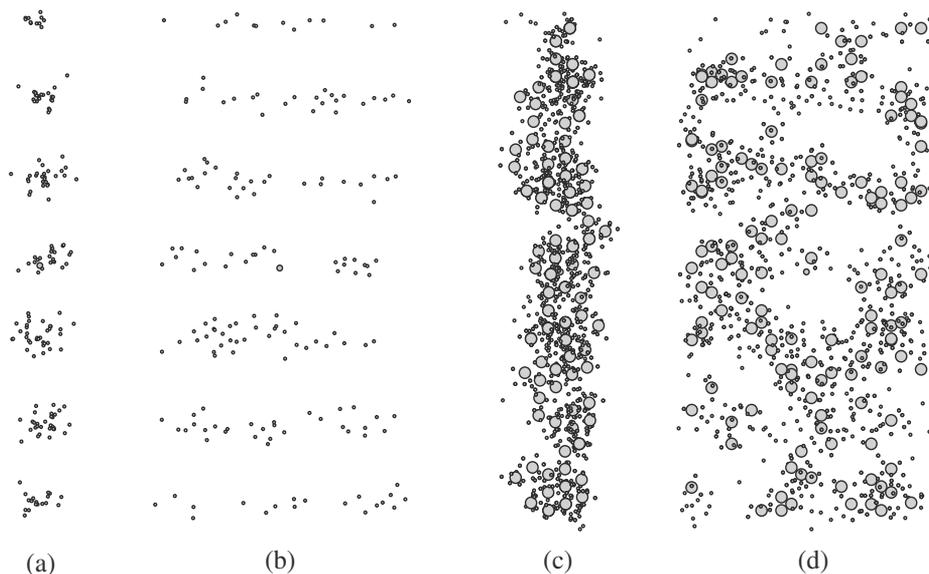


Fig. 3. A configuration taken from the Monte Carlo simulation of hydrogen segregation at the grain boundary shown on Fig. 1. (a) is a projection along the tilt axis of the grain boundary, like Fig. 1, (b) is a side view. The box contains 6 E units which lead to this “ladder” structure: H segregates along the tilt axis in the E units. The results are obtained by Monte Carlo simulations with H exchanges and displacements of Al and H particles. The total number of H particles is fixed, no metal vacancies are allowed. The temperature is 400K and the average bulk concentration is 300ppm atomic. The average grain boundary concentration is 15%, i.e. 15H atoms are present for 100 Al atoms in the region of half thickness  $1 a_0$  around the grain boundary plane. (c) and (d) are similar views, but in the case where the total H content is increased and that vacancy creation is introduced.

H-H repulsions, it never happens, unless a vacancy is present. At each step in the code, care is taken not to favor one representation of a state with respect to other ones. So, if no vacancies are present, and if we want to calculate the occupancy of one interstitial site, the sum should be made over all the different representations (Each physical T site, is represented by 4 different occupancies. All these should be summed to give the true occupancy of the site). In the case where some cells are multi-occupied (which could happen, in principle, only at very high H concentrations), the degeneracy of the labeling should be taken into account in the partition function. For the moment, we have not treated this case.

#### 4.3 Intergranular vacancy concentration in the presence of hydrogen

When the Monte Carlo simulation is run with a large enough total content of H and low temperature, the segregation at the grain boundary is important. That is, in realistic conditions like  $C_H = 100ppm$  and  $T=300K$ , the local concentration is of the order of 3 H for 10 metal atoms, which is large. If the concentration is increased, the level of H is large enough to disrupt the order in the boundary, as can be seen from the spreading of the peaks of the density profile perpendicular to the boundary. By doing so, the grain boundary tries to accommodate the large concentration by modifying its structure. This is a good case to test

the possibility to form large concentrations of vacancies. A first run was made at  $T=400\text{K}$  and  $C_H > 300\text{ppm}$ , with insertion/deletions, cluster moves, exchanges and displacements. The acceptance rate is too low to reach convergence with this brute force approach. Nevertheless, it can be seen (Fig. 3 c and d) that the grain boundary is enriched by a large number of vacancies. This confirms that superabundant vacancies should be present and well localized in the grain boundaries, since, in these temperature and concentration conditions they are not stable in the bulk (with this interaction energy  $-0.3\text{eV}$  and formation energy of  $0.7\text{eV}$ ). It is a strong indication that vacancies should be considered when large concentrations of solutes, in particular interstitial solutes, segregate. Technically, it is crucial to understand why the cluster moves have such a low acceptance rate in the grain boundary. To start with, we put aside the complex co-segregation of vacancies and solute interstitials and come back to the problem of simulating intergranular vacancies and especially their ordering at high concentration and low temperature.

## 5. Hybrid Monte Carlo

The strong relaxations induced in the grain boundary by the presence of vacancies severely reduce the efficiency of the vacancy exchange moves described above. In practice, if the simulation is started with all the vacancies in the bulk, the simple exchange moves can succeed in sampling the different configurations (i.e. the way the vacancies are distributed on the nodes), because the acceptance rate is not vanishingly small (see Fig. 2). On some GB sites, the relaxations bring the neighbors close to the node of the vacant site. Then, if the vacancy is exchanged with an atom, the relaxed neighbors and the new occupant overlap. The energy change becomes largely positive and the move is rejected. It becomes crucial when these sites are also those where the formation energy of the vacancy is the smallest, i.e. when the sites are the most occupied by vacancies. In this case it is impossible to get the proper concentrations and clustering of the vacancies.

Figure 4 quantifies this effect in the case of an exchange of a vacancy from one of such sites to another, geometrically equivalent one. A Monte Carlo simulation is run with exchange and displacement moves. The variation of the system energy during the exchange move is recorded. Its distribution is given by the curve labeled "simple X" (X stands for exchange). It is wide and centered around  $7\text{eV}$ . Note that no moves gave energy variations lower than  $1\text{eV}$ . It means that the acceptance rate, at low temperature, is vanishingly small. As a consequence, the system is trapped with vacancies occupying always the same sites. In particular, the geometrically equivalent sites are never visited. To solve the problem of the overlap, a first idea is to couple the exchange with displacement moves. Monte Carlo simulations are run first without vacancies and then with a single vacancy at a fixed location. The displacements of the particles are collected and the distributions are calculated for each node. They are usually gaussians, centered at, or close to, the node position (by construction of the lattice). Of course, the distributions are translated and distorted for the neighbors of the vacancy due to the large relaxations. A translation towards the vacancy is observed, but on a very limited number of sites. The idea is to measure the range of these relaxations and identify which neighbors are mostly affected. It defines a sphere of interaction around a vacancy (and, to insure reversibility, around the future vacancy node) where all the particles are attributed a new displacement after the exchange. Consider the initial vacancy, surrounded by relaxed neighbors. A particle is brought into the vacant site. A new displacement is taken according to the distribution measured without vacancies, for all the neighbors in the sphere (respectively for all the neighbors of the vacancy at its new location, according the distributions in the

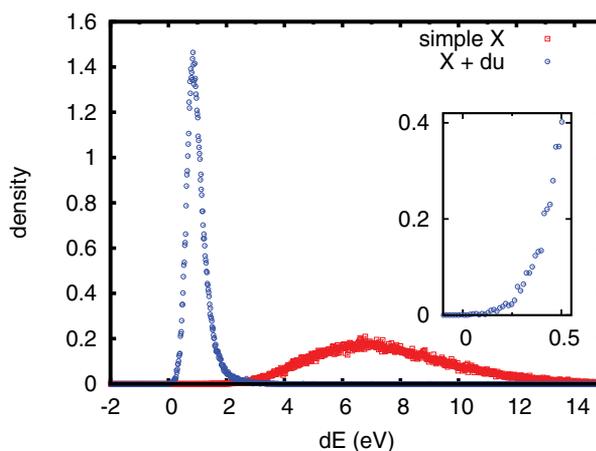


Fig. 4. Energy distribution for an exchange move between a vacancy with strong relaxations and a particle. Curve labeled “simple X” represents the variation of the energy obtained during the exchange alone, with the overlap between the relaxed neighbors and the new particle. Curve “X + u” is obtained by combining the exchange with a displacement move which reduces the overlap. The inserted graph shows that moves with a low energy variation are rare (no moves with negative energy variation were found).

presence of vacancies). The metropolis criterion is modified to respect detailed balance. The new energy distribution for this “compound move” is shown on Fig. 4 (curve “X + u”). It is re-centered closer to the low energies, but still the probability that the energy variation is lower than 0.5eV is very small (see the inserted graph on Fig. 4). This was tested systematically by introducing more and more neighbors in the “interaction sphere”, without success. The problem is that the new displacements are selected, at random, independently. There is no notion of collective arrangement of the neighbors.

There are several methods which bias the choice of the displacement of the atoms by taking into account the direction of the force and, usually, a random term (Allen & Tildesley (1991)) is added: “smart Monte Carlo” (Rosky et al. (1978)), “dynamic Monte Carlo” (Kotelyanskii & Suter (1992)), “force-bias Monte Carlo” (Cao & Berne (1990))... The idea is to be guided by a “cheap” dynamics of the system to choose the trial positions. Faster convergence is achieved in the cases where concerted movements are necessary. In the problem described above, we want to make a large jump in configuration space by transporting, arbitrarily, the vacancy from one location to another. At the same time, we want to have both the initial and the final surroundings of the vacancy elastically relaxed and thermally excited (Uhlherr & Theodorou (2006)) like they should be in equilibrium. In the next section, the exchange move is combined with Molecular Dynamics (MD) to obtain this effect.

### 5.1 Algorithm

Hybrid Monte Carlo (Duane et al. (1987)) (HMC) is a method which uses MD to generate a trial state and accepts it with the usual Metropolis rule. It has been adapted to condensed matter (Mehlig et al. (1992)) and applied in dense liquids (Desgranges & Delhommelle (2008)), for example. It requires some modifications to be adapted to the vacancy problem that are detailed now.

How can the exchange and the HMC moves be combined? The algorithm must be reversible and must satisfy detailed balance. The exchange alone is reversible. The HMC move is done by first, in the old state characterized by the occupancy of the nodes and the displacements of the particles (with some tolerance), taking random velocities for the particles according to the Maxwell distribution corresponding to the temperature  $T$ . The integration of the motion is then performed with a time reversible, area preserving algorithm (the volume of phase space is preserved). For simplicity, the iso kinetic (Gauss) velocity verlet algorithm from Tuckerman is used (Tuckerman (2010)), so there is no need to consider the kinetic energy term in the energy balance. The new state is obtained after  $2n$ MD integration steps. The reverse move is performed by selecting exactly the opposite velocities. The integration of  $2n$ MD steps brings back the system to the original configuration.

If the MD steps are done after the exchange, the algorithm is not reversible. If one wants to use time reversibility, the exchange must be inserted in the middle of the MD trajectory. Then the compound move is:

- select velocities according to the Maxwell distribution
- perform  $n$ MD time reversible MD steps
- project the configuration on the lattice
- select a vacancy and a particle
- exchange them
- perform  $n$ MD time reversible MD steps
- project the configuration on the lattice
- calculate the energy variation and accept the new state according to the Metropolis rule

The “projection” is done by looping over the mobile particles and testing if they still belong to their original Voronoï cell. If not, the new Voronoï cell is found. If it is occupied by a particle, the move is not considered (not rejected, in the sense that it is not considered as a valid attempt). If it is occupied by a vacancy, the book keeping is made to exchange the occupancies of the sites and update the displacements.

Performing MD steps on the whole system is not necessary and prohibitively expensive. To speed up the calculations, only a limited number of particles are considered mobile. The compound move has two parameters:  $n$ MD, the number of MD steps done before and after the exchange and a sphere radius  $r_{MD}$  to define the mobile particles. Sometimes, a vacancy is brought by the exchange on a GB site where it is not stable. Then, even if the MD trajectory is short, the vacancy diffuses. This induces some practical complications for choosing, reversibly, the vacancy and the particle for the exchange and calculate the probabilities that appear in the detailed balance.

## 5.2 Detailed balance

Let’s take a closer look at the detailed balance condition in the case of a move from the bulk to the grain boundary. As mentioned earlier, it is not necessary to use the hybrid move for all the exchange moves, but only for some specific grain boundary sites. The sites can be decomposed in two categories: the “in” sites and the “out” sites. “in” means “in the region of the GB where the relaxations are particularly strong”. The different moves can be  $in \rightarrow out$ ,  $in \rightarrow in$ ,  $out \rightarrow in$  and  $out \rightarrow out$ . The  $out \rightarrow out$  transition is done with simple exchanges. Let’s focus on the  $out \rightarrow in$  move. For the discussion of the detail balance, it is called the “direct” move and the inverse move is called the “return” move.

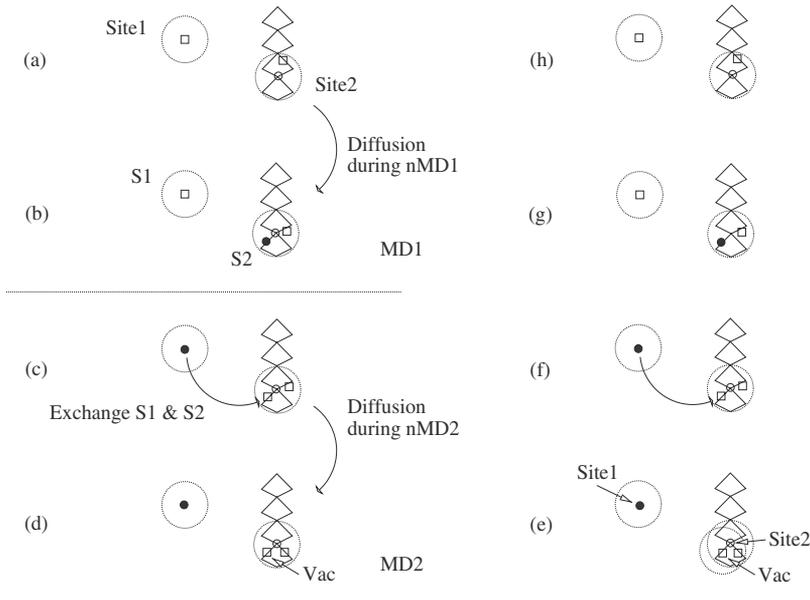


Fig. 5. Schematic representation of a hybrid Monte Carlo move bringing a bulk vacancy in the grain boundary (a to d) and the reverse move (e to h). The squares represent vacancies and the small circles, particles.

A vacant site is taken at random, with probability  $1/nVac$ , where  $nVac$  is the number of vacancies in the system. If the site is “out” (resp. “in”) of the GB region, a trial move to bring the vacancy “in” (resp. “out”) is constructed. Site1 is the vacant site selected. In the case of the *out*  $\rightarrow$  *in* move it is in the bulk (Fig. 5a). Then another site is taken at random in the list of the “in” sites, with probability  $1/nGBsites$  ( $nGBsites$  is the number of sites considered “in”). This site is called Site2 (Fig. 5a). The list of mobile particles is constructed from the list of the *nodes* contained in two spheres of radius  $r_{MD}$  around Site1 and Site2 (Fig. 5a). Then the velocities are taken randomly according to the Maxwell distribution, and  $nMD$  Molecular Dynamics steps are performed. Diffusion events occur during this relaxation, due to exchanges between mobile particles and vacancies (Fig. 5b). Note that, for preserving reversibility, if a mobile particle is exchanged with a vacancy not contained in the list of mobile sites, the move is abandoned. The particles are projected on the lattice, i.e. only one particle is affected to each Voronoï cell. If this is not possible, the move is abandoned. If the vacancy on Site1 diffuses, the move is abandoned because it is not possible to construct the return move. Otherwise, it is labeled S1. A site, labeled S2, is taken at random among the “in” sites contained in the list of mobile sites, with a probability  $1/nbGBm$  ( $m$  means “mobile”). The simple exchange move is performed between S1 and S2.  $nMD$  Molecular Dynamics steps are performed. The final configuration is projected on the lattice and the move is abandoned if it is not possible. A final check is performed on the vacancies which is necessary to guarantee reversibility. In particular, at the end of this procedure, at least one vacancy should be present in the sphere of radius  $r_{MD}$  around Site2. It is labeled Vac (Fig. 5d). The motivation for this condition appears below, during the construction of the return move.

For the return, a way must be found to trace back Site1 and Site2, in the new configuration, and construct exactly the same list of mobile particles. A vacancy is taken at random. The

probability to select Vac (Fig. 5e) is  $1/nVac$ . It is "in". An "in" site is taken at random in the sphere of radius  $r_{MD}$  around Vac. The probability that the choice gives Site2 is  $1/nbGBm'$ , where  $nbGBm'$  is the number of "in" sites in the sphere around Vac. This number has to be calculated at the end of the direct move (in d), because it enters detailed balance. But to be in the situation of doing such calculation, there must be at least one vacancy in the sphere around Site2 (that we called Vac by anticipation). For reversibility, this condition is imposed in the last step of the direct move (in d), as mentioned above.

An "out" site is taken at random. The probability to find Site1 is  $1/n_{out}$ , where  $n_{out}$  is the number of sites not in the list of "in" sites. The mobile particles are created from the nodes included in two spheres of radius  $r_{MD}$  around Site1 and Site2. They are the same than those of the direct move. The velocities are taken exactly as the opposite of the velocities at the end of the direct move, with the same probability as in the direct move. The trajectory is reversed during nMD steps. A vacancy is taken at random, among the "in" sites included in the list of mobile sites. The number of these vacancies is  $nVacm$  (m stands for "mobile"). The probability to choose the same vacancy as the one appearing after the exchange in the direct move is  $1/nVacm$ . This probability has to be evaluated during the direct move after the exchange. The exchange is performed. nMD steps are made to recover the initial configuration (a).

The condition imposing that a vacancy remains in the sphere around Site2, at the end of the direct move, is not a restrictive one because: first, only mobile particles can diffuse, and therefore, the vacancy always stays in the mobile list; second, it is rare that the two spheres join. Another condition, is that no particles are allowed to leave from the mobile list, by an exchange with a vacancy on a node outside the mobile list. Otherwise, this particle will not be mobile during the inverse move.

The difficulty, when designing this algorithm, lies on the fact that the vacancies diffuse quite strongly during the relaxation. This is essentially because the exchanges force the occupancy of GB sites near the most favourable sites. A simple algorithm which abandons the move as soon as it detects that a vacancy leaves its original site during MD, gives a high abandon rate, of the order of 70%. It is a large waste of computation time and also a waste of information on the system because these fast diffusion paths are interesting to learn about the diffusion mechanism, in a statistical sense.

The "underlying matrix"  $\alpha$  is not symmetric. Detailed balance is:

$$\frac{1}{Z} \exp(-\beta H_0) \times \alpha_{out \rightarrow in} acc_{out \rightarrow in} = \frac{1}{Z} \exp(-\beta H_n) \times \alpha_{in \rightarrow out} acc_{in \rightarrow out} \quad (9)$$

$$\alpha_{out \rightarrow in} = 1/nVac \times 1/nGBsites \times 1/nbGBm \quad (10)$$

$$\alpha_{in \rightarrow out} = 1/nVac \times 1/nbGBm' \times 1/n_{out} \times 1/nVacm \quad (11)$$

or

$$\frac{acc_{out \rightarrow in}}{acc_{in \rightarrow out}} = \exp(-\beta(H_n - H_0 - kT \ln(\alpha_{in \rightarrow out}/\alpha_{out \rightarrow in}))) \quad (12)$$

The metropolis criterion is:

- If  $H_n - H_0 - kT \ln(\alpha_{in \rightarrow out}/\alpha_{out \rightarrow in}) \leq 0$  the move is accepted
- If  $r = H_n - H_0 - kT \ln(\alpha_{in \rightarrow out}/\alpha_{out \rightarrow in}) > 0$ , a random number (rand) is taken between 0 and 1. If  $rand < \exp(-\beta r)$ , the move is accepted, otherwise it is rejected and the system stays in the old state (i.e. the same state is repeated twice in the Markov chain).

A similar algorithm is used for the  $in \rightarrow out$  moves, and a simpler one for the  $in \rightarrow in$  moves with a symmetric "underlying matrix".

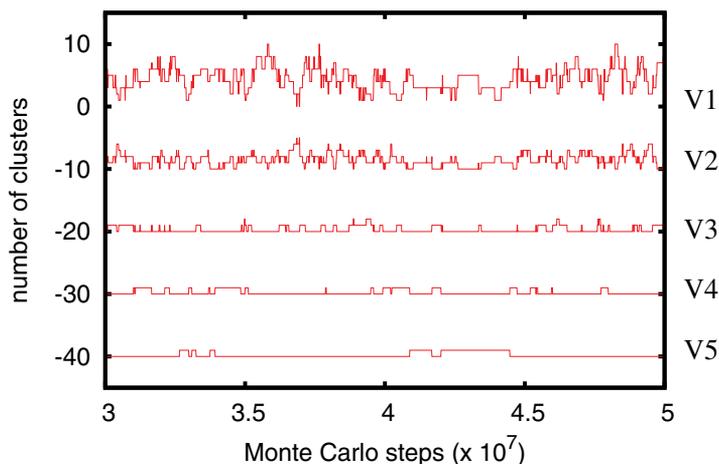


Fig. 6. Fluctuation of the number of vacancy clusters containing from 1 (V1) to 5 (V5) vacancies, as a function of the number of Monte Carlo steps. The total number of vacancies is 10, the total number of accessible sites is 144. Temperature is 200K. Each curve is translated down by  $-10$  with respect to the previous one, for clarity.

### 5.3 A first application to vacancies at a grain boundary

This methodology is tested (Tanguy (n.d.)) by simulating the equilibrium of a fixed number of vacancies, in Al, in a box containing the symmetrical tilt grain boundary  $\Sigma 33(554)[110]$  (Vamvakopoulos & Tanguy (2009)). The dimensions are such that three structural units (SU) (each one composed of two twin SU and two specific units of this family of tilt boundary, called E) are taken in the direction perpendicular to the tilt axis (Y) and 12 SU along it (Z direction). The calculation of the formation energy of the isolated vacancy shows (Vamvakopoulos & Tanguy (2009)) that the E unit contains 2 atomic sites where this value is significantly lower than in the bulk. The presence of a vacancy in one of these sites induces large relaxations with all the problems of low acceptance rate for the exchanges discussed above. All these sites are gathered in a list, the list of “in” sites. The total number of “in” sites is 144. The total number of sites is of the order of 17000. Periodic boundary conditions are applied in the Y and Z directions, while rigid borders are imposed in the direction perpendicular to the interface (X direction). This is done to impose the lattice parameter of the perfect crystal at a reasonable distance from the interface.

The Monte Carlo simulation is composed of two types of moves:

- individual displacements of the particles in the whole system, as is commonly done to simulate the canonical ensemble
- compound hybrid Monte Carlo-exchange moves (HMC-X).

The proportions are 0.9995 displacements for 0.0005 HMC-X moves. Every 10000 microstep, a picture of the box, giving the location of each vacancy, is taken. Each of these images is post treated. At low temperature, the structure of the GB is such that the vacancies tend to align along the tilt axis, with some kinks. This ordering is roughly quantified by identifying the number of clusters of each size, from 1 to the total number of vacancies. A cluster is defined like a chain of first neighbors (it doesn't have to be a straight line).

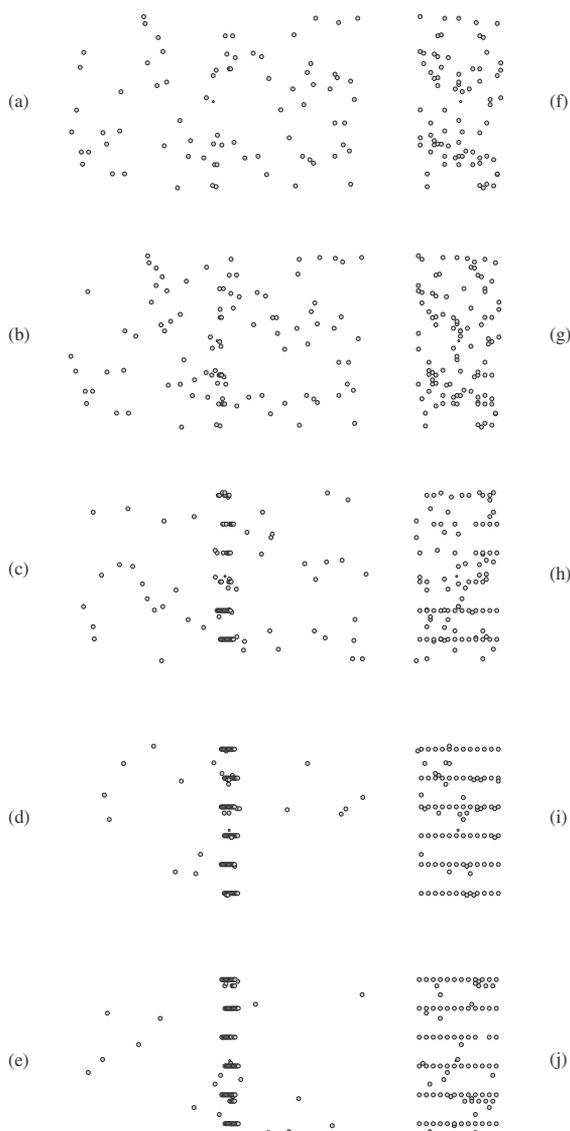


Fig. 7. A sequence of states, during equilibration of the system: (a) to (e) are front views (slightly tilted), (f) to (j) are the corresponding side views. The circles represent vacancies. The atoms are not shown.

Before trying to get some physical informations out of the simulations, it is important to evaluate its efficiency, i.e. to determine which range of thermodynamics variables (number of vacancies, temperature) that can be used as parameters for the Monte Carlo simulation and get a reasonable convergence, i.e. a meaningful approximate of the cluster distribution. Of course, this information should be obtained in a reasonable computer time (less than a month, on a single processor). A first test is made by taking 10 vacancies and distributing them

among 144 “in” sites. Temperatures between 200K and 500K were used. Figure 6 shows the fluctuation of the number of clusters during the simulation (after some equilibration). It seems that a single long chain of vacancy is unstable with respect to shorter segments, even at the low temperature of 200K, in spite of a strong effective pair interaction (0.3eV (Vamvakopoulos & Tanguy (2009))). This graph and a visual control show that the system is not trapped and that the clusters are moved around the grain boundary as they should be when only short range order is established. The quantitative measure of the concentrations (Tanguy (n.d.)), as a function of temperature, still requires a check of the dependency on the number of MD steps in HMC-X and the radius of relaxation (these parameters where  $n_{MD} = 200$  and  $r_{MD} = 1.6a_0$ , where  $a_0$  is the lattice parameter of Al). The method looks promising: the acceptance rate for HMC-X moves between geometrically equivalent sites is in between 30% and 40%, with a weak temperature dependence. The drawback is its high computational cost, which means that a tuning should be done to fix the proportions of regular exchange moves (without any relaxations) and HMC-X.

The last example (Fig. 7), is the equilibration, at T=200K, of a simulation containing 100 vacancies. An equal proportion of  $in \rightarrow in$  and  $in \rightarrow out$  moves are proposed. The vacancies are randomly distributed in the initial configuration (Fig. 7a and f) and gradually lead to the saturation of the lines, parallel to the tilt axis, composed of the sites where the formation energy is the lowest. Some more complex clusters are found that will be described later (Tanguy (n.d.)).

## 6. Conclusion

This paper shows the different steps in the development of a Monte Carlo simulation which gives, at the atomic scale, the equilibrium segregation of interstitial solutes and the vacancy concentration at pre-existing crystalline defects. The focus is on grain boundaries, but it can be adapted to dislocations. The flexibility in the design of the Monte Carlo moves, in particular unphysical moves like exchanges, cluster moves and the mixing with Molecular Dynamics, enables the sampling of a complex phase space of large dimensions and reveals details of the structure of the vacancy clusters that can not be guessed intuitively from the GB structure. Some work remains to be done to simulate the co-segregation of interstitial solutes and vacancies, with enough statistics to guarantee that the ordering of the  $VH_n$  clusters is properly sampled.

This work is supported by ANR blan2006 Hinter and blan2010 EcHyDNA.

## 7. References

- Adams, D. J. (n.d.). *Mol. Phys.* 29: 307. .
- Allen, M. P. & Tildesley, D. J. (1991). *Computer Simulation of Liquids*, Oxford University Press, New York.
- Bagchi, K., Andersen, H. C. & Swope, W. C. (1996). *Phys. Rev. Lett.* 76: 255.
- Bugeat, J. P., Chami, A. C. & Ligeon, E. (1976). *Phys. Rev. Lett.* 58A: 127.
- Cao, J. & Berne, B. J. (1990). *J. Chem. Phys.* 92: 1980.
- Dai, J., Kanter, J. M., Kapur, S. S., Seider, W. D. & Sinno, T. (2005). *Phys. Rev. B* 72: 134102.
- Denkiewicz, M. J., Hoagland, R. G. & Hirth, J. P. (2008). *Phys. Rev. Lett.* 100: 136102.
- Desgranges, C. & Delhommelle, J. (2008). *Phys. Rev. B* 77: 054201.
- Duane, S., Kennedy, A. D., Pendleton, B. J. & Roweth, D. (1987). *Phys. Lett. B* 195: 216.
- Dünweg, B. & Landau, D. P. (1993). *Phys. Rev. B* 48: 14182.

- Fukai, Y. (1993). *The Metal Hydrogen System*, Springer.
- Fukai, Y. & Ōkuma, N. (1994). *Phys. Rev. Lett.* 73: 1640.
- Ganchenkova, M. G., Borodin, V. A. & Nieminen, R. M. (2009). *Phys. Rev. B* 79: 134101.
- Gunaydin, H., Barabash, S. V., Houk, K. N. & Ozolinš, V. (2008). *Phys. Rev. Lett.* 101: 075901.
- Henkelmann, G. & Jónsson, H. (2001). *J. Chem. Phys.* 115: 9657.
- Kotelyanskii, M. J. & Suter, U. W. (1992). *J. Chem. Phys.* 96: 5383.
- Landau, D. & Binder, K. (2000). *A Guide to Monte Carlo Simulation in Statistical Physics*, Cambridge University Press.
- Liu, Y.-L., Zhang, Y., Zhou, H.-B. & Lu, G.-H. (2009). *Phys. Rev. B* 79: 172103.
- Mehlig, B., Heermann, D. W. & Forrest, B. M. (1992). *Phys. Rev. B* 45: 679.
- Mezei, M. (n.d.). *Mol. Phys.* 40: 901.
- Nørskov, J. K. & Besenbacher, F. (1987). *J. Less-Common Met.* 130: 475.
- Rosky, P. J., Doll, D. J. & Friedman, H. L. (1978). *J. Chem. Phys.* 65: 4628.
- Rowley, L. A., Nicholson, D. & Parsonage, N. G. (n.d.). *Mol. Phys.* 31: 365.
- Soisson, F., Barbu, A. & Martin, G. (1996). *acta mater.* 44: 3789.
- Sørensen, M. R., Mishin, Y. & Voter, A. F. (2000). *Phys. Rev. B* 62: 3658.
- Sørensen, M. R. & Voter, A. F. (2000). *J. Chem. Phys.* 9599: 112.
- Suzuki, A. & Mishin, Y. (2005). *J. Mater. Sci.* 40: 3155.
- Swope, W. C. & Andersen, H. C. (1992). *Phys. Rev. A* 46: 4539.
- Swope, W. C. & Andersen, H. C. (1996). *J. Chem. Phys.* 102: 2851.
- Tanguy, D. (n.d.). *in preparation*.
- Tanguy, D. & Mareschal, M. (2005). *Phys. Rev. B* 72: 174116.
- Tateyama, Y. & Ohno, T. (2003). *Phys. Rev. B* 174105: 67.
- Tuckerman, M. (2010). *Statistical Mechanics, Theory and Molecular Simulations*, Oxford University Press, New York.
- Uhlherr, A. & Theodorou, D. N. (2006). *J. Chem. Phys.* 125: 084107.
- Valleau, J. P. & Cohen, L. K. (1980). *J. Chem. Phys.* 72: 5935.
- Vamvakopoulos, E. & Tanguy, D. (2009). *Phys. Rev. B* 79: 094116.
- van Meegen, W. & Snook, I. (1980). *J. Chem. Phys.* 73: 4656.
- Voter, A. F., Montalenti, F. & Germann, T. C. (2001). *Ann. Rev. Mater. Res.* 32: 321.
- Wang, J.-C., Pal, S. & Fichthorn, K. A. (2001). *Phys. Rev. B* 63: 085403.
- Wolverton, C., Ozolinš, V. & Asta, M. (2004). *Phys. Rev. B* 69: 144109.

# Frequency-Dependent Monte Carlo Simulations of Phonon Transport in Nanostructures

Qing Hao and Gang Chen  
*Massachusetts Institute of Technology*  
USA

## 1. Introduction

It is now widely recognized that phonon transport in micro/nanostructures is strongly affected by interfaces and boundaries, which can lead to a significantly reduced thermal conductivity (Cahill et al., 2003; Chen, 2005; McConnell & Goodson, 2005). On one hand, the reduced thermal conductivity can significantly impede heat spreading in microelectronic and photonic devices, creating challenges to thermal management (Garimella et al., 2008). On the other hand, such size effects have been exploited to develop nanostructured bulk materials with better thermoelectric (TE) performance (Hsu et al., 2004; Poudel et al., 2008; Ma et al., 2008; Wang et al., 2008; Joshi et al., 2008; Yang et al., 2009; Zhu et al., 2009). The effectiveness of a TE material is ultimately determined by its dimensionless figure of merit (ZT), defined as  $ZT = S^2\sigma T / k$ , where  $S$ ,  $\sigma$ ,  $k$ ,  $T$  represent Seebeck coefficient, electrical conductivity, thermal conductivity, and absolute temperature, respectively (Goldsmid, 1964). In nanostructured bulk materials, nanostructured interfaces can be designed to strongly scatter phonons but only slightly affect the charge carrier transport. This approach leads to a significant reduction in the lattice thermal conductivity  $k_L$ , and in some cases a simultaneous increase of the power factor  $S^2\sigma$  (Minnich et al., 2009a), resulting in ZT improvements over their bulk counterparts. Unlimited to traditional TE materials, the nanostructuring approach may also yield a high ZT in materials that were previously unsuitable for TE applications due to their high thermal conductivities. This concept was demonstrated in nanostructured bulk silicon, where a ZT around 0.7 was achieved at around 1200 K (Bux et al., 2009). With bulk-like electrical properties and significantly reduced thermal conductivity, ZT=0.6 at room temperature was reported for rough silicon nanowires (Hochbaum, 2008). Low thermal conductivities were also found in silicon nanomesh structures, with electrical conductivities preserved from bulk silicon in the high-doping range (Yu et al., 2010). Along this line, ZT around 0.4 was measured at room temperature for silicon membranes with high density of nanoscopic holes (Tang et al., 2010). Despite these promising results, very little theoretical work has been conducted to understand the phonon transport in nanostructured bulk materials, which hinders the prediction of their potential ZT improvements. There have been efforts applying the Boltzmann Transport Equation (BTE) to periodic 2D structures (Chung & Kaviany, 2000; Yang & Chen, 2004; Yang et al., 2005; Prasher, 2006; Miyazaki et al., 2006; Pattamattaa & Madnia, 2009), but with the gray medium approximation, i.e., a frequency-independent phonon mean free path (MFP). For arbitrary structures, Monte Carlo (MC) simulations are

more favorable to solving the BTE. Only considering the boundary scattering of phonons, MC simulations were first conducted to understand the 1D thermal-conductance measurements on polished single crystals of pure silicon at low temperatures (Klitsner et al., 1988). Taking the internal collisional processes of phonons into account, Peterson designed a MC scheme based on the gray-medium approximation to study 1D heat transfer problems (Peterson, 1994). More recently, the MC technique was applied to 2D nanowire composites (Jeng et al., 2008; Tian & Yang, 2007) and 3D nanoparticle composites (Jeng et al., 2008). The work by Jeng et al. was the first phonon MC simulations for a composite system. Nevertheless, the gray-medium approximation employed in both studies could lead to a significant underestimation of phonon size effects inside micro- to nano-structured bulk materials. For example, experimental results have suggested strong phonon size effects in micro-porous silicon films (Song & Chen, 2004), which can be explained only if frequency-dependent phonon MFPs are considered (Hao et al., 2009). With advancements of theoretical and numerical tools (McGaughey & Kaviani, 2005; Broido et al., 2007; Lindsay & Broido, 2008; Henry & Chen, 2008; Turney et al., 2009), more accurate information on phonon MFPs for some materials are becoming available. Such advancements call for the development of numerical tools that can include the frequency dependence of phonon transport (Narumanchi et al., 2006).

In this chapter, a MC simulation technique considering frequency-dependent phonon MFPs is introduced to investigate the phonon transport in various periodic structures. A novel boundary condition based on the periodic heat flux with a constant virtual wall temperature is developed for the studied periodic structures. This allows us to calculate the thermal conductivity of a periodic structure with a single period as the computational domain. In the literature, frequency-dependent phonon MC simulations were first performed on solid thin films (Mazumder & Majumdar, 2001). However, the suggested treatment for three-phonon scattering, including normal process (N process) and Umklapp process (U process), violated energy conservation before and after the three-phonon scattering events and thus led to errors in simulations. To correct this, three-phonon scattering treatment respecting the energy conservation was proposed later (Lacroix et al., 2005; Chen et al., 2005; Hao et al., 2009). Unlimited to thin films, the MC simulations were also carried out for nanowires (Chen et al., 2005; Lacroix et al., 2006; Randrianalisoa & Baillis, 2008) or 1D transient phonon transport in bulk materials (Lacroix et al., 2005). Among all listed studies, the N process and U process were mostly treated together by a combined scattering rate except for that Chen et al. adopted a genetic algorithm to satisfy both energy and momentum conservation for the N process and energy conservation for the U process (Chen et al., 2005). As an extension, our work introduced in this chapter is the first attempt to apply the MC method to complicated geometries with inclusion of frequency-dependent phonon scattering. The simulation code is used to compute the lattice thermal conductivities of 2D porous silicon with periodically aligned pores (Hao et al., 2009) and 3D silicon nanoparticle composites (Hao et al., 2010). The latter one is essentially nano-grained bulk silicon.

For lightly doped silicon, frequency-dependent MC simulations show that a large thermal conductivity reduction can be observed even in micro-porous structures, an effect that is not expected if an average phonon MFP is used. This indicates that phonon size effects in lightly doped nanostructured bulk silicon can be significantly underestimated if frequency-dependent phonon MFPs are not included into the model.

For heavily doped silicon nanocomposites used for thermoelectrics, phonons are strongly scattered by charge carriers and point defects inside the grains. In analyzing the

experimental data of n-type silicon nanocomposites with grain sizes around 200 nm, we find that grain interface scattering of phonons is negligible and much smaller grain sizes are required to obtain remarkable phonon size effects. Based on the parameters extracted from fitting experimental results, we predict a ZT around 1.0 at 1173 K as grain sizes are reduced to 10 nm. This ZT value is comparable to traditional SiGe bulk alloys and can be further improved by optimizing electrical properties. Compared to other TE materials using exotic and expensive elements, pure silicon nanocomposites have significant advantages for commercialization in terms of cost and material abundance.

Unlimited to 2D porous silicon and 3D silicon nanocomposites, MC simulations can also be applied to other materials with various structures and geometries. The work demonstrated in this book chapter sets up the platform for lattice thermal conductivity predictions across a wide range of length scales.

## 2. Basic simulation scheme

In a MC simulation, phonons bundles are first drawn and distributed randomly across the computational domain. Each bundle represents a number of phonons with similar properties. Their initial states (velocity, frequency, branch, and traveling direction) are generated by a random sampling approach based on the equilibrium phonon spectrum (Mazumder & Majumdar, 2001; Chen et al., 2005), which will be discussed later. With their individual velocities and traveling directions, phonons are allowed to move and may experience various scattering events during their movement. At each time step, whether a phonon will get scattered and thus change its state is determined by a random number and the individual scattering probabilities. By tracking a large number of phonons and averaging the results over a long period of time, statistically the MC simulations will approach the BTE solutions after convergence.

As an overview, the procedure of MC simulations is briefly described here. We still follow the schematic process flow given in the previous work (Jeng et al., 2008), but the gray-medium approximation is replaced by the frequency-dependent model. In the simulation, the computational domain is divided into many spatial bins, also called subcells. Because it is not feasible to simulate a large number of phonons, phonons are grouped into bundles (with  $W$  phonons per bundle) to save computer memory. States of phonons inside each bundle are identical, i.e., they share the same angular frequency  $\omega$ , traveling direction  $\vec{k}$ , polarization  $p$ , and group velocity  $V_{g,p}(\omega)$ , where the subscript  $p$  indicates the polarization. In this chapter, only the longitudinal acoustic (LA) branch and two transverse acoustic (TA) branches will be considered because the optical branches contribute little to the thermal conductivities directly (Chen, 2005). The potential effect of optical phonon scattering on the acoustic phonon MFPs (Lindsay & Broido, 2008) is included via the molecular dynamics simulation results themselves (Henry & Chen, 2008), from which phonon MFPs are obtained for our MC simulations.

At the beginning of each simulation, phonon bundles are generated inside the computational domain according to the initial temperatures assigned to individual subcells. The states of created phonons are randomly sampled based on the equilibrium phonon spectrum  $\langle n \rangle D(\omega)$  (Mazumder & Majumdar, 2001; Chen et al., 2005), where  $D(\omega)$  is the density of states for phonons,  $\langle n \rangle$  is the Bose-Einstein distribution at the current subcell temperature  $T$ , defined as

$$\langle n \rangle = \frac{1}{\exp\left(\frac{\hbar\omega}{k_B T}\right) - 1}. \quad (1)$$

For a subcell with volume  $V_{sub}$ , the number of created phonon bundles is  $V_{sub} \sum_{p=1}^3 \int_0^{\omega_{p,max}} \langle n \rangle D(\omega) d\omega / W$ , in which  $\omega_{p,max}$  is the maximum phonon frequency for branch  $p$ , and  $W$  is again the phonon bundle size. Initially, the created phonon bundles are randomly distributed spatially inside each subcell. In all simulations,  $W$  is chosen so that the total number of initialized phonon bundles inside the whole domain is less than  $1.5 \times 10^7$ . Although smaller  $W$  is always preferred for less fluctuations in simulation results, the computing speed can be extremely slow with more phonon bundles in the domain. To define the frequency of a phonon bundle,  $\omega_0$ , a random number  $R$  ( $0 \leq R \leq 1$ ) is generated and  $\omega_0$  value should satisfy

$$R = \frac{\sum_{p=1}^3 \int_0^{\omega_0} \langle n \rangle D(\omega) d\omega}{\sum_{p=1}^3 \int_0^{\omega_{p,max}} \langle n \rangle D(\omega) d\omega}. \quad (2)$$

The exact polarization of the phonon is determined by another random number between zero and unity. It indicates the LA branch if the number is less than the ratio

$$P = \frac{[\langle n \rangle D(\omega)]_{LA}}{2[\langle n \rangle D(\omega)]_{TA} + [\langle n \rangle D(\omega)]_{LA}} \Big|_{\omega=\omega_0}, \quad (3)$$

in which  $[\langle n \rangle D(\omega)]_{LA}$  and  $[\langle n \rangle D(\omega)]_{TA}$  represent the product  $\langle n \rangle D(\omega)$  for the LA branch and TA branch, respectively. Otherwise, the phonon bundle belongs to a TA branch. After the polarization and angular frequency of a phonon bundle are both determined, its group velocity  $V_{g,p}(\omega)$  can be obtained from the phonon dispersion curve. For the 3D simulation, the traveling direction of a phonon is generated by two random numbers,  $R_1$  and  $R_2$  ( $0 \leq R_{1,2} \leq 1$ ). The unit vector of the traveling direction is

$$\vec{k} = \begin{pmatrix} \sin\theta \cos\psi \\ \sin\theta \sin\psi \\ \cos\theta \end{pmatrix}, \quad (4)$$

where the polar angle  $\theta$  satisfies  $\cos\theta = 2R_1 - 1$ , and the azimuthal angle  $\psi$  is determined by  $\psi = 2\pi R_2$ .

Within a time step, each phonon bundle travels with its own group velocity, which is  $V_{g,L}(\omega)$  for the longitudinal mode and  $V_{g,T}(\omega)$  for the transverse mode. To achieve good spatial resolutions, the time step  $\Delta t$  is chosen so that the maximum travel distance of a phonon bundle within  $\Delta t$ ,  $V_{max}\Delta t$ , is smaller than the subcell dimension. Consequently, phonon bundles will generally take a few time steps to travel out of a subcell, which enables us to better capture the phonon movement. During their travels, the phonons may

encounter the interfaces or the domain boundaries, and will change their trajectories. The interface transport treatment will be discussed in the following subsections. After the phonon movement, the total phonon energy inside each subcell is calculated and divided by the subcell volume to get the phonon energy density, defined as

$$E = \sum_{p=1}^3 \int_0^{\omega_{p,\max}} \hbar\omega \langle n \rangle D(\omega) d\omega, \quad (5)$$

in which the pseudo-temperature  $\tilde{T}$  of a subcell appears in the term

$$\langle n \rangle = \frac{1}{\exp\left(\frac{\hbar\omega}{k_B \tilde{T}}\right) - 1}, \quad (6)$$

and  $\tilde{T}$  can be computed by the numerical inversion of Eq. (5) (Mazumder & Majumdar, 2001; Chen et al., 2005; Hao et al., 2009). The above expression works for phonons in equilibrium, but becomes invalid when we deal with transport processes. Within an infinitesimal time step  $\Delta t$ , the density of scattered phonons can be expressed as

$$N_s = \sum_{p=1}^3 \int_0^{\omega_{p,\max}} \frac{\Delta t}{\tau(\omega)} \langle n \rangle D(\omega) d\omega, \quad (7)$$

where  $\tau(\omega)$  is the scattering relaxation time, the weight  $\Delta t / \tau(\omega)$  is the probability of being scattered, the subscript  $s$  indicates scattered phonons. Similarly, the energy density of scattered phonons is

$$E_s = \sum_{p=1}^3 \int_0^{\omega_{p,\max}} \hbar\omega \frac{\Delta t \langle n \rangle}{\tau(\omega)} D(\omega) d\omega, \quad (8)$$

instead of the energy density  $E$  in Eq. (5). For a fixed time step  $\Delta t$ , we can always use

$$\frac{E_s}{\Delta t} = \sum_{p=1}^3 \int_0^{\omega_{p,\max}} \hbar\omega \frac{\langle n \rangle}{\tau(\omega)} D(\omega) d\omega \quad (9)$$

to evaluate the temperature of scattered phonons, denoted as  $T_s$ . In thermal equilibrium,  $T_s$  is equal to the subcell temperature  $\tilde{T}$  though  $T_s$  is only sampled among all scattered phonons inside a subcell.

The above approach is consistent with the results from the BTE. Under the relaxation time approximation, the phonon BTE can be written as (Goodson et al., 1997)

$$\frac{\partial f}{\partial t} + V_{g,p}(\omega) \vec{k} \cdot \nabla_{\vec{r}} f = -\frac{f - \langle n \rangle}{\tau(\omega)}, \quad (10)$$

where  $\vec{k}$  is the unit vector of the traveling direction,  $\langle n \rangle$  is the Bose-Einstein distribution,  $\tau(\omega)$  is the scattering relaxation time, vector  $\vec{r}$  is the phonon position,  $f$  is the phonon distribution function,  $V_{g,p}(\omega)$  is the group velocity, and subscript  $p$  indicates the polarization. We multiply both sides of Eq. (10) by  $\hbar\omega D(\omega)$  and integrate with respect to phonon angular frequency  $\omega$ . The summation over the three acoustic branches yields

$$\sum_{p=1}^3 \int_0^{\omega_{p,\max}} \frac{\partial f}{\partial t} \hbar \omega D(\omega) d\omega + \sum_{p=1}^3 \int_0^{\omega_{p,\max}} \hbar \omega D(\omega) V_{g,p}(\omega) \vec{k} \cdot \nabla_{\vec{r}} f d\omega = - \sum_{p=1}^3 \int_0^{\omega_{p,\max}} \hbar \omega D(\omega) \frac{f - \langle n \rangle}{\tau(\omega)} d\omega \quad (11)$$

On the other hand, phonon energy conservation requires

$$\frac{\partial u}{\partial t} + \nabla_{\vec{r}} \cdot \vec{q} = 0, \quad (12)$$

in which  $u = \sum_{p=1}^3 \int_0^{\omega_{p,\max}} f \hbar \omega D(\omega) d\omega$  is the phonon energy density, and the vector  $\vec{q} =$

$\sum_{p=1}^3 \int_0^{\omega_{p,\max}} \hbar \omega D(\omega) V_{g,p}(\omega) \vec{k} f d\omega$  is the heat flux. For the equilibrium situation,  $u$  is equalized

to  $E$  defined in Eq. (5) to get the pseudo-temperature  $\tilde{T}$  appearing in  $\langle n \rangle$ . However, this is not the case when we are only concerned with the scattered phonons. In Eq. (12), it can be observed that

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial t} \left[ \sum_{p=1}^3 \int_0^{\omega_{p,\max}} f \hbar \omega D(\omega) d\omega \right] = \sum_{p=1}^3 \int_0^{\omega_{p,\max}} \frac{\partial f}{\partial t} \hbar \omega D(\omega) d\omega, \quad (13)$$

and

$$\nabla_{\vec{r}} \cdot \vec{q} = \sum_{p=1}^3 \int_0^{\omega_{p,\max}} \hbar \omega D(\omega) V_{g,p}(\omega) \vec{k} \cdot \nabla_{\vec{r}} f d\omega. \quad (14)$$

Comparing Eqs. (11) and (12), we obtain

$$\sum_{p=1}^3 \int_0^{\omega_{p,\max}} \hbar \omega D(\omega) \frac{f - \langle n \rangle}{\tau(\omega)} d\omega = 0,$$

or

$$\sum_{p=1}^3 \int_0^{\omega_{p,\max}} \hbar \omega D(\omega) \frac{f}{\tau(\omega)} d\omega = \sum_{p=1}^3 \int_0^{\omega_{p,\max}} \hbar \omega D(\omega) \frac{\langle n \rangle}{\tau(\omega)} d\omega, \quad (15)$$

which indicates that for scattered phonons the numerical inversion process to get  $\tilde{T}_s$  must be associated with  $\sum_{p=1}^3 \int_0^{\omega_{p,\max}} \hbar \omega D(\omega) \frac{\langle n \rangle}{\tau(\omega)} d\omega$  instead of the energy density  $E$  defined in

Eq. (5). Under the gray-medium approximation,  $\tau(\omega)$  is a constant for all the phonons at the same temperature, and Eq. (15) will reduce to the case defined by Eq. (5). For the frequency-dependent model, however, cautions must be taken and Eq. (15) must be used to define the temperature of scattered phonons.

The phonons may also experience internal scatterings during their travels. These scatterings are processed at the end of each time step. The scattering probabilities, used to determine

the occurrence of internal scattering events, are first computed based on the updated subcell temperature  $\tilde{T}$ , phonon branch, and phonon frequency. The selection rule and detailed treatment of scattered phonons will be given later. Basically the simulation repeats the following procedures after the phonon initialization: phonon movement with possible interface or boundary scatterings, subcell temperature update with the new spatial distribution of phonon bundles, internal scattering treatment inside each subcell. Convergence is achieved when the calculated temperature profile no longer changes with time. After the convergence, the thermal conductivity of the studied structure can be derived from the domain dimension, the heat flowing through the domain, and the temperature difference across the domain.

### 3. Phonon scattering treatment

In MC simulations, phonon bundles may experience internal scatterings (by impurities, charge carriers, or other phonons) or interface scattering, where charge carrier scattering of phonons are normally neglected for lightly doped samples (Mazumder & Majumdar, 2001). The detailed treatments are discussed below.

#### 3.1 Internal scattering treatment

For internal scatterings, we assume that charge carrier scattering and impurity scattering only randomize the phonon bundle traveling direction, leaving the phonon frequency and velocity unchanged. In contrast, the phonon-phonon scattering, including the N process and U process, will reset all phonon states (its frequency, branch, velocity, and traveling direction) according to the current subcell temperature. This will help establish the thermal equilibrium between the scattered phonon bundle and the local temperature.

The phonon overall lifetime  $\tau_T$  is calculated using the Matthiessen's rule,  $\tau_T^{-1}(\omega) = \tau_I^{-1}(\omega) + \tau_N^{-1}(\omega) + \tau_U^{-1}(\omega) + \tau_{E(H)}^{-1}(\omega)$ , where  $\tau_I$ ,  $\tau_{E(H)}$ ,  $\tau_N$ ,  $\tau_U$  are the relaxation times of impurity scattering, charge carrier scattering ( $\tau_E$  for electrons,  $\tau_H$  for holes), for N processes, and U processes, respectively. For lightly doped samples, the term  $\tau_E^{-1}(\omega)$  or  $\tau_H^{-1}(\omega)$  can be neglected. In Callaway's model (Callaway, 1959), N processes were treated differently from other scattering mechanisms, by adding an associated correction term to the thermal conductivity. However, the refinement work by Holland (Holland, 1963) argued that this correction term would possibly cause more errors and N processes should be treated in exactly the same way as other scattering mechanisms. In this work, we follow Holland's treatment and N processes are simply a scattering mechanism added to other mechanisms. Additionally, the N process and the U process are not differentiated, and a combined relaxation time  $\tau_{NU}$ , defined as  $\tau_{NU}^{-1}(\omega) = \tau_N^{-1}(\omega) + \tau_U^{-1}(\omega)$ , will be used. The probability for a phonon bundle to be scattered is given by  $P(\omega) = 1 - \exp(-\Delta t / \tau_T(\omega))$  (Mazumder & Majumdar, 2001; Hao et al., 2009). For an infinitesimal  $\Delta t$ ,  $P(\omega)$  simply becomes the weight  $\Delta t / \tau_T(\omega)$  in Eqs. (7) and (8). At the end of every time step, a random number  $R_s$  ( $0 \leq R_s \leq 1$ ) is generated for each phonon bundle at a frequency  $\omega_0$  and compared with  $P(\omega_0)$ . The phonon bundle is scattered if  $R_s$  is less than  $P(\omega_0)$ . If a phonon bundle is scattered, another random number  $R_{NU}$  will be generated and compared with  $P_{NU}(\omega_0) = \tau_{NU}^{-1}(\omega_0) / \tau_T^{-1}(\omega_0)$ . The phonon will have phonon-phonon scattering if  $R_{NU}$  is less than  $P_{NU}(\omega_0)$ . Otherwise, it will be scattered by either charge carriers or impurities, which will only randomize its traveling direction.

Phonon-phonon scattering will reset all the states of a phonon bundle, which follows similar procedures as described by Eqs. (2) to (4). However, the spectrum  $\langle n \rangle D(\omega)$  in Eqs. (2) and (3) should be replaced by  $\langle n \rangle D(\omega) / \tau_{NU}(\omega)$  introduced in Eqs. (7) to (9). Because  $1 / \tau_{NU}(\omega)$  monotonically increases with  $\omega$ , most phonons scattered are from the high-frequency portion of the phonon spectrum. As a result, at a fixed temperature the averaged frequency of the phonons scattered by other phonons,

$$\langle \omega \rangle_s = \frac{\sum_{p=1}^3 \int_0^{\omega_{p,\max}} \frac{1}{\tau_{NU}(\omega)} \hbar \omega \langle n \rangle D(\omega) d\omega}{\sum_{p=1}^3 \int_0^{\omega_{p,\max}} \frac{1}{\tau_{NU}(\omega)} \hbar \langle n \rangle D(\omega) d\omega}, \quad (16)$$

is higher than the averaged frequency of all the phonons at the same temperature, given as

$$\langle \omega \rangle_{eq} = \frac{\sum_{p=1}^3 \int_0^{\omega_{p,\max}} \hbar \omega \langle n \rangle D(\omega) d\omega}{\sum_{p=1}^3 \int_0^{\omega_{p,\max}} \hbar \langle n \rangle D(\omega) d\omega}. \quad (17)$$

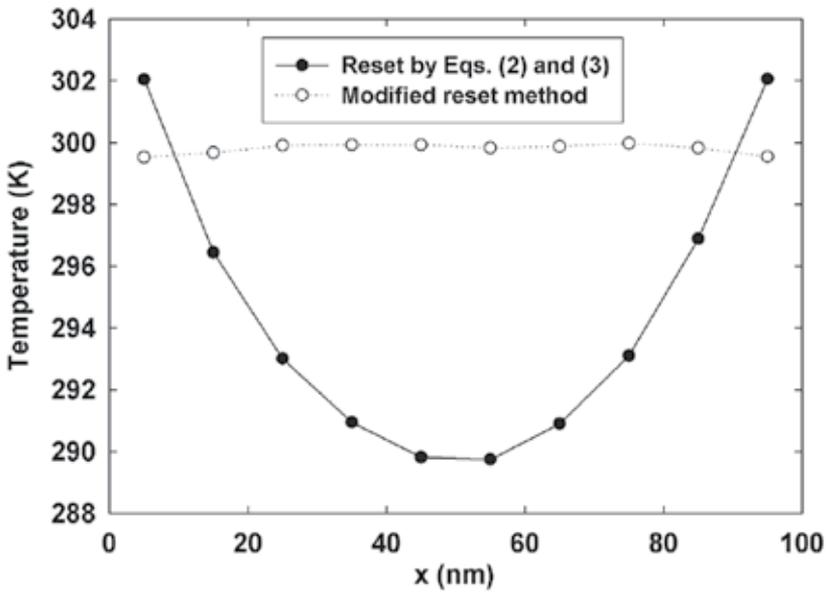


Fig. 1. Calibration of the reset methods for scattered phonons (Hao et al., 2009). The slight asymmetry and undulation of the temperature profiles are attributed to the numerical errors of MC simulations. The 1D computational domain is 100 nm in length and its both ends are fixed at 300 K. The reset method based on Eqs. (2) and (3) always yields an abnormal temperature profile, which can be resolved by replacing the spectrum  $\langle n \rangle D(\omega)$  with  $\langle n \rangle D(\omega) / \tau_{NU}(\omega)$  in Eqs. (2) and (3). Here the time step is fixed at 1 ps and the temperature of each subcell is averaged over 0.5 ns.

In the literature, Eqs. (2) and (3) were used to reset the states of a scattered phonon bundle (Mazumder & Majumdar, 2001). The phonon bundles experiencing phonon-phonon scattering are assigned frequencies averaged at  $\langle \omega \rangle_{eq}$ . We find that this treatment will result in a final phonon spectrum distorted from the real spectrum and will yield abnormal temperature profiles when the temperature difference across the domain is small. A simple test can be conducted with the equilibrium situation in which the 1D computational domain has both ends fixed at 300 K (isothermal wall boundary condition in Mazumder & Majumdar, 2001). To simplify, the model used in this test only considers the U processes and impurity scattering (Dames & Chen, 2005). The corresponding scattering rates are:  $1/\tau_U(\omega) = B_1 \nu^2 T \exp(-B_2/T)$  (U process), where  $\nu$  is the phonon frequency,  $B_1 = 3.0 \times 10^{-19}$  s/K,  $B_2 = 210$  K;  $1/\tau_I(\omega) = A\omega^4$  (impurity scattering), with  $A = 2.4 \times 10^{-45}$  s<sup>3</sup>. The Debye model is used for the phonon dispersion, with  $\omega_{LA,max} = \omega_{TA,max} = 7.06 \times 10^{13}$  rad/Hz, and the lattice constant  $a = 5.5$  Å. Due to the distorted phonon spectrum from the equilibrium one, the resulting temperature profile always shows a valley in the middle (solid circles in Fig. 1), with a significantly increased number of phonon bundles from equilibrium. This problem is resolved when we use the spectrum  $\langle n \rangle D(\omega) / \tau_{NU}(\omega)$  instead of  $\langle n \rangle D(\omega)$  in Eqs. (2) and (3) to reset the scattered phonon states (empty circles in Fig. 1). This treatment is consistent with Eq. (15), obtained from the BTE under the relaxation time approximation.

After processing all internal scatterings inside one single subcell, phonons will be either randomly created according to the spectrum  $\langle n \rangle D(\omega)$  in Eqs. (2) and (3) or deleted from the domain. Our purpose is to assure the subcell energy conservation before and after the scatterings. The energy imbalance inside a subcell is controlled to be less than  $\hbar\omega_{TA,max} / 2$ .

### 3.2 Interface scattering treatment

For micro- and nanostructured materials, both the interface scattering and internal scattering can contribute to the thermal resistance of the material. We assume phonons are diffusely scattered at the interfaces. When a phonon bundle encounters an interface during its travel, it will be diffusely transmitted or reflected. For phonons incident from side 1 of the interface, this is determined by a random number  $P_r$  ( $0 \leq P_r \leq 1$ ) and the interface transmissivity  $T_{12}$  from side 1 toward side 2 (Tian & Yang, 2007; Jeng et al., 2008; Hao et al., 2009; Hao et al., 2010). The phonon bundle will be transmitted to side 2 if  $P_r$  is less than  $T_{12}$ . Otherwise, it will be reflected back to side 1. The reflected or transmitted phonon bundles are assigned a new traveling direction, with all other phonon states (velocity, frequency, and branch) unchanged, and hence phonon mode conversion is not included. For rectangular subcells, the direction reassignment within the anticipated semi-sphere for either reflection or transmission is simple. Equation (4) is again used to generate a direction vector  $\vec{k}$  and then the sign of one particular component in  $\vec{k}$  will be specified. It should be noted that  $\theta$  should be determined by  $\sin^2 \theta = R_1$  instead of  $\cos \theta = 2R_1 - 1$  in this case (Jeng et al., 2008). Suppose a phonon bundle moving in the positive x direction is transmitted across a y-z plane interface. It will have a positive x component in its new traveling vector  $\vec{k}$ . If it is reflected, the x component will be negative. Similarly, we can treat reflection and transmission on x-y and x-z plane interfaces. After the interface scattering, the phonon bundle will continue its movement with the remaining drift time to finish the current time step.

#### 4. Boundary condition

For a periodic structure, the boundary condition used for the numerical solution of the 2D BTE equation (Yang & Chen, 2004) can be extended to the 3D MC simulation. With a  $L_x \times L_y \times L_z$  rectangular computational domain, heat is assumed to flow in the positive  $x$  direction. Our chosen unit cell is symmetric in both  $y$  and  $z$  directions. For such a symmetric structure, phonon bundles hitting the four side walls of the domain are specularly reflected (Tian and Yang, 2007; Jeng et al., 2008; Hao et al., 2009; Hao et al., 2010). If the unit cell structure is asymmetric in the  $y$  or  $z$  directions, we can use periodic boundary conditions for the corresponding side walls. In this case, phonon bundles exiting from the domain on one side wall will re-enter the domain from the same interception position on the opposite side wall, and continue their travels. In the  $x$  direction, the essence of the boundary condition is that on both ends of the simulation domain, which we will call the hot and cold walls, the distortions of the distribution functions from the equilibrium distribution are periodic (shown in Fig. 2). This can be written as

$$f_2^-(\theta, \psi, \omega, y, z) - \langle n(y, z) \rangle_2 = f_1^-(\theta, \psi, \omega, y, z) - \langle n(y, z) \rangle_1, \quad (18)$$

$$f_2^+(\theta, \psi, \omega, y, z) - \langle n(y, z) \rangle_2 = f_1^+(\theta, \psi, \omega, y, z) - \langle n(y, z) \rangle_1, \quad (19)$$

in which  $f$  is the distribution function,  $\theta$  is the polar angle,  $\psi$  is the azimuthal angle, the subscript "2" donates the cold wall ( $x = L_x$ ) and "1" represents the hot wall ( $x = 0$ ), the superscripts "+" and "-" represent the distribution in the positive and negative  $x$  directions. Here  $\langle n(y, z) \rangle_1$  and  $\langle n(y, z) \rangle_2$  are evaluated at the corresponding local wall temperatures. They are not periodic since their difference is the driving force for heat flow along the  $x$  direction.

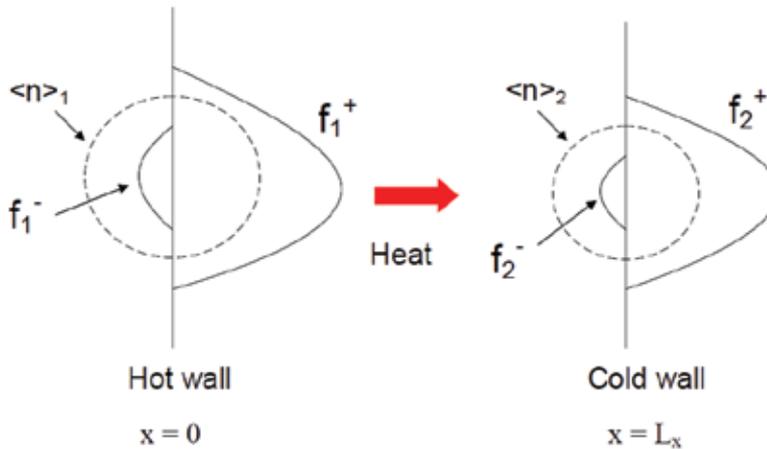


Fig. 2. Distribution functions on the domain boundaries described by the periodic heat flux boundary condition. The hot wall and cold wall positions are located at  $(0, y, z)$  and  $(L_x, y, z)$ , respectively. The equilibrium distributions,  $\langle n(y, z) \rangle_1$  and  $\langle n(y, z) \rangle_2$ , are both isotropic. With the  $x$ -direction heat flow, the distribution function  $f$  is anisotropic and distorted from  $\langle n(y, z) \rangle$ .

Equation (18) means when a phonon bundle, following the distribution function  $f_1^-(\theta, \psi, \omega, y, z)$ , encounters the hot wall during its movement, it will be “absorbed” by the wall. At the same time, a new phonon bundle, following the distribution function  $f_2^-(\theta, \psi, \omega, y, z)$ , will be emitted from the cold wall to maintain the heat flow through the domain. Similarly, Eq. (19) relates the distribution functions of phonon bundles “absorbed” by the cold wall and phonon bundles emitted from the hot wall. In our simulations, the absorbed phonons are recorded during tracking the phonon movement in each time step. The states of the emitted phonons are determined by Eqs. (18) and (19), as described below.

#### 4.1 Cold wall emission

In Eq. (18),  $f_2^-(\theta, \psi, \omega, y, z)$  represents the distributions of phonons entering the simulation domain through the cold wall, which we call cold wall emitted phonons. The distribution of these emitted phonons is determined by  $f_1^-(\theta, \psi, \omega, y, z)$ , the distributions of phonons leaving the domain from the hot wall, or absorbed phonons. In the  $y$ - $z$  plane, a fixed cross section area  $A$  is assigned for all subcells. Denoting unit vector  $\vec{n}$  as the normal of the wall,  $\vec{k}$  as the phonon traveling direction, we multiply  $(V_{g,p}(\omega)\vec{k}\cdot\vec{n})D(\omega)A\Delta t$  to both sides of Eq. (18) and integrate with respect to  $\omega$ , and  $\psi$ ,  $\theta$  for  $2\pi$  solid angle. Summation over the three acoustic branches and rearranging the equation yields

$$N_2^-(y, z) = N_1^-(y, z) - \frac{1}{4W} \sum_{p=1}^3 \int_0^{\omega_{p,\max}} V_{g,p}(\omega) D(\omega) A \Delta t \left( \langle n(y, z) \rangle_1 - \langle n(y, z) \rangle_2 \right) d\omega, \quad (20)$$

where  $W$  is the number of phonons in each bundle,  $N_1^-(y, z)$  is the number of locally absorbed phonon bundles on the hot wall,  $N_2^-(y, z)$  is the number of phonon bundles locally emitted by the cold wall. Similarly, multiplying  $\hbar\omega(V_{g,p}(\omega)\vec{k}\cdot\vec{n})D(\omega)A\Delta t$  to both sides of Eq. (18) and conducting similar procedures yields

$$Q_2^-(y, z) = Q_1^-(y, z) - \frac{1}{4W} \sum_{p=1}^3 \int_0^{\omega_{p,\max}} \hbar\omega V_{g,p}(\omega) D(\omega) A \Delta t \left( \langle n(y, z) \rangle_1 - \langle n(y, z) \rangle_2 \right) d\omega, \quad (21)$$

in which  $Q_2^-(y, z)$  is the total phonon bundle energy locally emitted from the cold wall for area  $A$ ,  $Q_1^-(y, z)$  is the total phonon bundle energy absorbed by the hot wall.

In our simulations, all the  $N_1^-(y, z)$  phonon bundles are deleted from the domain but their states are separately stored as a pool for the cold wall emission. The phonon states include its frequency, velocity, location intercepting the wall, the remaining drift time after hitting the wall, and its traveling direction. To obtain the  $N_2^-(y, z)$  phonon bundles, we need to delete

$$N_{del} = \frac{1}{4W} \sum_{p=1}^3 \int_0^{\omega_{p,\max}} V_{g,p}(\omega) D(\omega) A \Delta t \left( \langle n(y, z) \rangle_1 - \langle n(y, z) \rangle_2 \right) d\omega, \quad (22)$$

phonon bundles from the pool consisting of  $N_1^-(y, z)$  phonon bundles. To do this,  $N_{del}$  phonons are first randomly generated according to the spectrum  $V_{g,p}(\omega)D(\omega)\left(\langle n(y, z) \rangle_1 - \langle n(y, z) \rangle_2\right)$ , which replaces  $\langle n \rangle D(\omega)$  in Eqs. (2) and (3) for phonon state determination in this situation. Among the  $N_1^-(y, z)$  phonon bundles, we delete  $N_{del}$

bundles that best match the frequencies and branches of the generated bundles. The states of these  $N_{del}$  bundles are saved for later usage, which will be clear when the hot wall emission is discussed. The phonons left in the pool constitute the  $N_2^-(y, z)$  phonons emitted from the cold wall.

#### 4.2 Hot wall emission

Similar to Eq. (20), the hot wall emission can be derived from Eq. (19) as

$$N_1^+(y, z) = N_2^+(y, z) + \frac{1}{4W} \sum_{p=1}^3 \int_0^{\omega_{p,max}} V_{g,p}(\omega) D(\omega) A \Delta t \left( \langle n(y, z) \rangle_1 - \langle n(y, z) \rangle_2 \right) d\omega, \quad (23)$$

in which  $N_2^+(y, z)$  phonon bundles are absorbed by the cold wall, and they are known for each time step. The previous  $N_{del}$  phonon bundles deleted from the cold wall emission pool, with their traveling directions reversed this time, will be added to the  $N_2^+(y, z)$  bundles and thus form the  $N_1^+(y, z)$  bundle pool for the hot wall emission.

#### 4.3 Periodic heat flux with a constant virtual wall temperature boundary condition

Based on the foregoing discussions, a novel periodic heat flux with a constant virtual wall temperature boundary condition can be introduced for periodic structures. We assign  $T_1(y, z) \equiv T_h$  to the hot wall and  $T_2(y, z) \equiv T_c$  to the cold wall of the domain. By combining such temperature setting with periodic boundary conditions used in Eqs. (18) and (19), the local heat flux  $q''$  is allowed to vary across each virtual wall but still hold its periodicity,  $q''(0, y, z) = q''(L_x, y, z)$ . It should be noted that our boundary condition is fundamentally different from the traditional isothermal wall boundary condition (Mazumder & Majumdar, 2001). In the latter case, the computational domain is sandwiched between two physical black walls. For a periodic structure, the isothermal wall boundary condition requires a computational domain consisting of many periods to eliminate the strong end effects. The exact thermal conductivity can be obtained only if the calculation results will no longer change with further increasing the number of periods inside the computational domain. In contrast, our boundary condition is specified on two "virtual" walls cut arbitrarily inside the material without affecting the real heat flow, as long as the virtual walls define an integer multiple of unit cells of the periodic structure. In principle, accurate thermal conductivities can be obtained using a single period as our computational domain. This will significantly reduce the computational cost. Our calculation validates this point.

### 5. Results and discussion

The previous sections cover the basic simulation procedure, the treatment of different phonon scatterings, and the employed boundary condition for a periodic structure. With these, we are ready to use frequency-dependent MC simulations to investigate the phonon transport in various periodic structures. My calculations here are focused on silicon-based materials, including 2D micro- to nano-porous silicon and 3D silicon nanocomposites.

#### 5.1 Employed models for different scattering mechanisms

To get dependable simulation results, it is necessary to use accurate frequency-dependent phonon lifetimes in silicon. In the literature, Holland's model (Holland, 1963) provided a

very good fit to the experimental data from 1.7 to 1300 K. However, in his work the real phonon dispersion was simplified as two linear sections for both TA and LA branches, resulting in abrupt changes of phonon group velocities across the threshold frequencies between two linear sections ( $\omega_1$  for two TA branches,  $\omega_4$  for the LA branch). In addition, the phonon-phonon scattering rates employed different expressions for the TA branches below and above  $\omega_1$ . As a result, the calculated phonon MFPs had a significant jump across the threshold frequencies. Although the first problem was fixed using the real phonon dispersion (Mazumder & Majumdar, 2001), the second issue still remained and led to a TA phonon MFPs jump from 9 nm to 179 nm at  $\omega_1$  for Si. Henry and Chen carried out molecular dynamics simulations to study the spectral dependence of phonon MFPs in silicon (Henry & Chen, 2008). The results of this past work are now incorporated into the MC simulations. Based on their calculations from 300 to 1000 K, the combined relaxation time for the N and U processes follows the expression  $\tau_{NU} = A_{NU}v^{-2}T^{-b}$ , where  $v$  (in Hz) is the phonon frequency,  $A_{NU} = 5.32 \times 10^{18} \text{ K}^{1.49}/\text{s}$ ,  $b = 1.49$  for LA phonons; and  $A_{NU} = 5.07 \times 10^{18} \text{ K}^{1.65}/\text{s}$ ,  $b = 1.65$  for TA phonons. To simplify the simulation, an isotropic phonon dispersion is assumed and the calculated (001) direction phonon dispersion is used to evaluate the frequency-dependent phonon group velocity and density of states. The whole phonon spectrum (from 0 to  $\omega_{LA,\max}$ ) is discretized into  $N_b$  equally spaced intervals. Numerical integrations with respect to the phonon frequency, such as the evaluation of Eq. (5), are all conducted by summing the integrals over all  $N_b$  intervals, in which the integrals are evaluated at the central frequency of each interval.

For doped bulk silicon, the impurity-phonon scattering should be considered in addition to the phonon-phonon scattering. The phonon scattering rate by impurities is expressed as (Klemens, 1955)

$$\tau_I^{-1}(\omega) = A\omega^4, \quad (24)$$

where the constant  $A = A_{\delta M} + A_{\delta R} + A_x$  (Asheghi et al., 2002). Here  $A_{\delta M}$ ,  $A_{\delta R}$ ,  $A_x$  correspond to the scattering due to the presence of impurity atoms, the induced strain by inserting impurity atoms into the lattice, and unintentional impurities and imperfections, respectively. The employed phonon dispersion suggests an average sound velocity as  $v_s = \left\{ \left[ 2v_{TA}^{-1} + v_{LA}^{-1} \right] / 3 \right\}^{-1} = 6127 \text{ m/s}$  and this  $v_s$  is used to compute  $A_{\delta M}$  and  $A_{\delta R}$ . All other parameters are unchanged from the previous work that uses  $v_s = 6400 \text{ m/s}$  instead (Asheghi et al., 2002).

For free charge carrier scatterings of phonons (only considered for heavily doped samples), the scattering rate is given as (Ziman, 1956 & 1957)

$$\tau_E^{-1}(\omega) = \frac{E_d^2 m^{*3} v_g}{4\pi \hbar^4 d} \frac{k_B T}{\frac{1}{2} m^* v_g^2} \times \left\{ \frac{\hbar\omega}{k_B T} - \ln \left[ \frac{1 + \exp \left[ \left( \frac{1}{2} m^* v_g^2 - E_F \right) / k_B T + \hbar^2 \omega^2 / 8 m^* v_g^2 k_B T + \hbar\omega / 2 k_B T \right]}{1 + \exp \left[ \left( \frac{1}{2} m^* v_g^2 - E_F \right) / k_B T + \hbar^2 \omega^2 / 8 m^* v_g^2 k_B T - \hbar\omega / 2 k_B T \right]} \right] \right\}, \quad (25)$$

where  $T$  is the absolute temperature,  $\hbar$  is the Planck constant,  $k_B$  is the Boltzmann constant,  $E_d$  is the acoustic deformation potential,  $m^*$  is the density of states effective mass,  $d$  is the density,  $v_g$  is the averaged phonon group velocity, and  $E_F$  is the Fermi energy calculated by the generalized Kane's model based on the carrier concentration (Kane, 1957; Kołodziejczak, 1961). Because the investigated silicon nanocomposites always have a carrier concentration much higher than the carrier concentration threshold  $3.0 \times 10^{18} \text{ cm}^{-3}$  for becoming metallic (Yamanouchi et al., 1967; Alexander & Hocomb, 1968), the scattering of phonons by bound electrons or holes will not be considered. Given the relaxation times of different mechanisms, the lattice thermal conductivity of bulk silicon can be calculated by (Chen, 2005; Holland, 1963)

$$k_L = \frac{1}{3} \sum_{p=1}^3 \int_0^{\omega_{p,\max}} C_p(\omega) V_{g,p}^2(\omega) \tau_{T,p}(\omega) d\omega, \quad (26)$$

where the subscript  $p$  indicates the polarization,  $C_p(\omega)$  is the spectral volumetric specific heat,  $V_{g,p}(\omega)$  is the phonon group velocity, and  $\tau_{T,p}(\omega)$  is the overall phonon relaxation time for branch  $p$  and frequency  $\omega$ .

## 5.2 Calibration for bulk silicon

The MC code developed is first calibrated with bulk silicon and only phonon-phonon scattering is considered. In this case, Eq. (26) predicts bulk  $k_L=146.5 \text{ W/m K}$  at 300 K. Because we only employ the phonon dispersion along the (001) direction for our calculations, this isotropic  $k_L$  value is lower than the prediction averaged in three major crystal directions by Henry and Chen (175 W/m K by the BTE approach, 166 W/m K by the Green-Kubo analysis), in which the phonon density of states, used in calculating  $C_p(\omega)$ , are summed over the Brillouin zone.

With  $N_b=2000$ , phonon bundle size  $W = 2$ ,  $T_h = 305 \text{ K}$  and  $T_c=295 \text{ K}$  are assigned to a rectangular computational domain that is 200 nm in length, and 10 nm  $\times$  10 nm for the cross section area. The time step  $\Delta t$  is fixed at 0.6 ps in this simulation. After 1 ns, the temperature profile (averaged over successive 800 time steps) converges and its variation in the following steps is within 0.5 % (Fig. 3). The linearity for the converged temperature profile is  $R^2=0.9987$ . The heat flow passing through the domain always fluctuates due to the randomness of phonon absorption by the hot and cold walls at different time steps. This heat flow is obtained by averaging the net heat flows across the two walls, which are  $Q_2^+ - Q_2^-$  for the cold wall and  $Q_1^+ - Q_1^-$  for the hot wall. After the convergence of the temperature profile, the heat flows at different time steps are further averaged over a period, with the highest and lowest 20 values excluded. To be accurate, the averaging period is chosen to be no shorter than the time for the temperature profile to converge. The lattice thermal conductivity can then be calculated with this averaged heat flow and the temperature difference across the domain. Averaging over a period of 16 ns yields a thermal conductivity of 142.9 W/m K, which is slightly lower than the expected 146.5 W/m K. In comparison, averaging the heat flows over any 2.4 ns period after the initial 1 ns gives thermal conductivities ranging from 141.8 to 145.4 W/m K, all with less than 4% relative error. At 300 K, our employed model predicts that around 53% of the lattice thermal conductivity is contributed by phonons with MFPs larger than the 200 nm domain length. The chances for these phonons to be scattered are relatively small but their contributions can still be included by averaging the results over a long period of time.

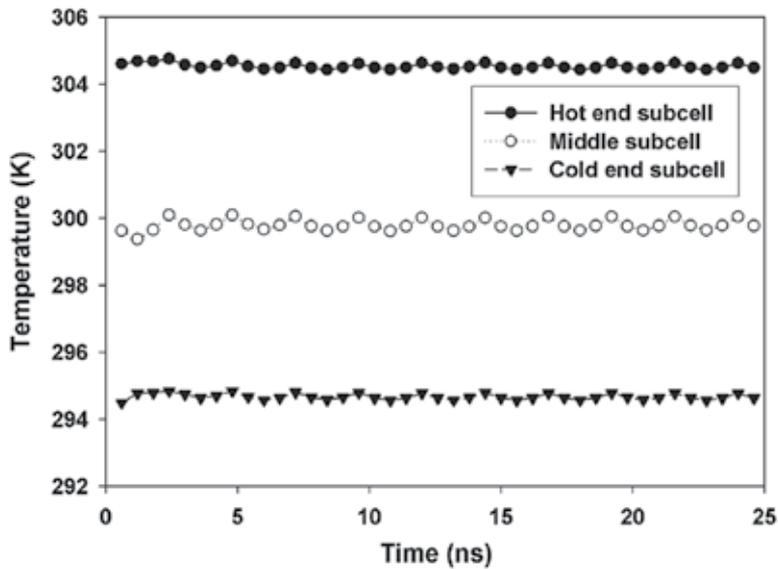


Fig. 3. The time history of subcell temperatures for a pure silicon computational domain (Hao et al., 2009). The  $200 \text{ nm} \times 10 \text{ nm} \times 10 \text{ nm}$  domain consists of 20 cubic subcells that are 10 nm in dimension. The subcell adjacent to the hot wall (hot end subcell), the one adjacent to the cold wall (cold end subcell), and the middle one are tracked here. Every temperature point is averaged over the previous 800 time steps, or 0.48 ns.

One important aspect of the statistical MC simulation is the signal-to-noise ratio, which is mainly reflected in the fluctuation of the heat flowing through the domain at different time steps. Such fluctuations can always be decreased by reducing the bundle size. However, this also requires a significantly larger computer memory that is not feasible in many cases. Averaging the heat flow over a longer period of time can be used to compensate the errors brought by large bundle sizes but requires more computational time. In practice, a balance between computation time and computer memory usage is required. The temperature difference between the cold wall and the hot wall can be increased to improve the signal-to-noise ratio, mainly by increasing the signal (averaged heat flow). The accuracy of the calculation will be slightly affected because in this case the thermal conductivity is an effective value averaged over a wider temperature range and does not correspond well to a particular temperature.

### 5.3 Two-dimensional porous silicon with aligned pores

Although it is one of the most important materials in the electronics industry, silicon is unsuitable for some applications because of its high thermal conductivity. Porous silicon, with its much lower thermal conductivity, could provide a simple solution to widen the usage of silicon (Yamamoto et al., 1999; Chung & Kaviani, 2000; Lee et al., 2007; Yu et al., 2010; Tang et al., 2010). Along this line, nanoscale porous structures are expected to introduce strong phonon size effects and further lower the thermal conductivity from the prediction based on the classical Fourier heat conduction theory. Surprisingly, experimental results showed that even microsize periodically arranged through-film pores would still yield notable phonon size effects in silicon thin films (Song & Chen, 2004). This reduction in

thermal conductivity cannot be explained using an averaged phonon MFP that is on the order of a hundred nanometer. To better understand the phonon size effects in porous silicon, we carry out frequency-dependent MC simulations on 2D porous silicon with periodically arranged square pores, in which the pore dimensions range from 5 nm to 2  $\mu\text{m}$ . It is assumed that phonons are all specularly reflected on the  $z$ -direction, i.e., the top and bottom surfaces of the film shown in Fig. 4a. This choice of the  $z$ -direction boundary condition converts the problem into a 2D case where the  $z$ -direction film thickness will not affect the results. The purpose of choosing this boundary condition is to reduce the computational load since a small  $z$ -direction dimension can be chosen (1 to 10 nm for all cases). Figure 4b presents the simulated film structure with the chosen computational domain, which is a square-shape period in the  $x$ - $y$  plane, with a square pore in its center. All phonons encountering the rough pore boundaries are reflected diffusely. The silicon film is assumed to be  $n$ -type with a doping level as  $5 \times 10^{15} \text{ cm}^{-3}$ . The electronic thermal conductivity is negligible ( $< 0.01 \text{ W/m K}$  from the Wiedemann-Franz Law) for this low doping level and the electron-phonon scattering can also be neglected. The constant  $A$  for impurity scattering is determined as  $9.3 \times 10^{-50} \text{ s}^3$ , with  $A_x$  as zero for slightly doped samples (Ashghi et al., 2002).

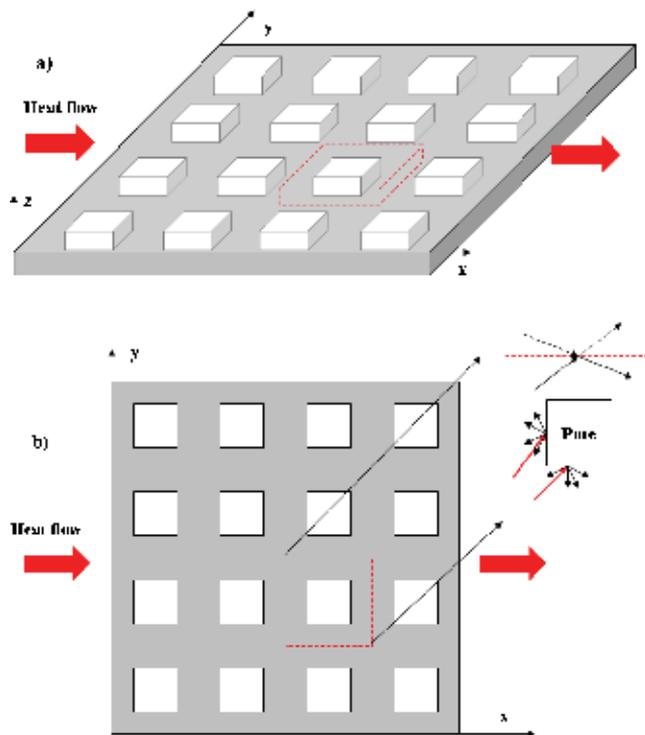


Fig. 4. a) Structure of the simulated porous silicon film. The chosen computation domain (marked by dashed line) is a single period with a square pore right in its center. b) Top view of the film, with details of the computational domain.

Based on the used parameters, the corresponding bulk lattice thermal conductivity is slightly reduced to  $145.9 \text{ W/m K}$  according to Eq. (26). In our simulations, the period of the

pores always equals twice the square pore size, resulting in a fixed porosity  $\Phi = 0.25$ . Without considering the phonon size effects, the Eucken model (Eucken, 1932 & 1933) from the Fourier classical heat conduction theory predicts

$$k_{Porous} / k_{Solid} = (1 - \Phi) / (1 + \Phi / 2), \quad (27)$$

for cubic pores, and the calculated  $k_{Porous}$  is 97.3 W/m·K.

A 10 K temperature difference is applied to the computational domain for cases with period sizes up to 1  $\mu\text{m}$ . To improve the signal-to-noise ratio, the temperature difference is increased to 20 K for larger period sizes. The normalized temperature contour for the 200-nm-period case is shown in Fig. 5a, with subcells chosen as 5 nm cubes. The normalization is performed as  $(T - T_{\min}) / (T_{\max} - T_{\min})$ , where  $T$  is the subcell temperature, and the subscripts max and min denote the maximum and minimum temperatures of subcells. Figure 5b presents the x-direction normalized temperature distribution at a few typical  $y$  locations. When the spacing between adjacent pores is smaller than the phonon MFPs, ballistic phonon transport becomes more important compared with the internal phonon scatterings inside silicon. Therefore, close to the left surface of the pore (facing the incoming heat flow) there will be a locally heated region because it receives hotter phonons directly from the “upstream” adjacent pore. Similarly, a locally cooled region will exist close to the back surface of the pore. In both cases, we will have a negative local temperature gradient compared with the main x-direction temperature gradient across the whole domain. These local “overshoots” do not violate the second law of thermodynamics because no local thermal equilibrium is established under the strong phonon ballistic transport (Yang & Chen, 2004; Jeng et al., 2008).

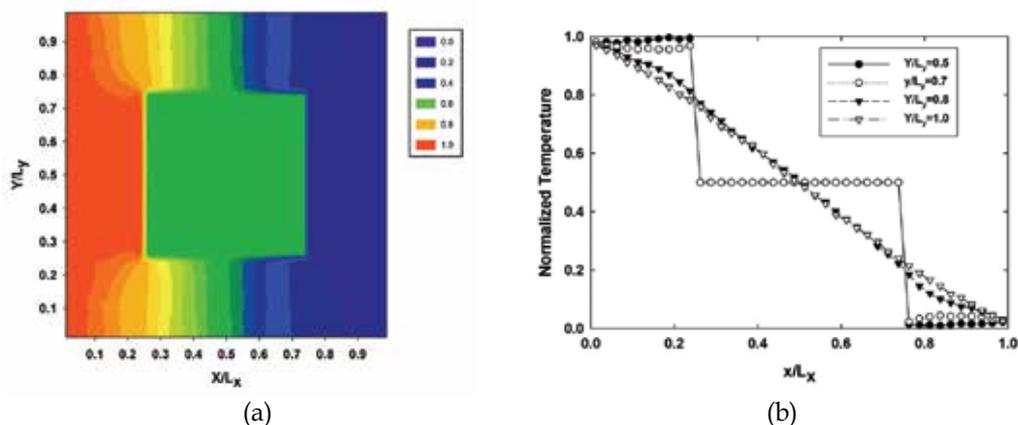


Fig. 5. (a) Normalized temperature contour for a 200 nm period (Hao et al., 2009). The empty pore region is assigned the average temperature of the domain. (b) Temperature distribution along the x direction for typical  $y$  locations.

The room-temperature in-plane lattice thermal conductivities, averaged over at least 4 ns after the convergence of the temperature profile, are plotted as a function of period size in comparison with the Eucken model prediction (Fig. 6). To clearly show the thermal conductivity reduction, all the values are normalized by the bulk thermal conductivity  $k_{Solid}$

(145.9 W/m·K). Remarkable phonon size effects can be observed even for a period size of 4  $\mu\text{m}$ , which is one magnitude larger than the averaged phonon MFP of 119 nm at 300 K. In comparison, solving the BTE based on the gray-medium approximation suggests  $k_{\text{Porous}} / k_{\text{Solid}}$  to be around 0.66 (shown in Fig. 6) for exactly the same 2D unit cell with a phonon Knudsen number of 0.1, where the Knudsen number is defined as the averaged phonon MFP divided by the period size (Miyazaki et al., 2006). Although isothermal wall boundary condition (Mazumder & Majumdar, 2001) is used instead of the periodic heat flux boundary condition to solve the BTE, this comparison clearly shows that phonon size effects cannot be accurately predicted without considering the frequency-dependent phonon MFPs.

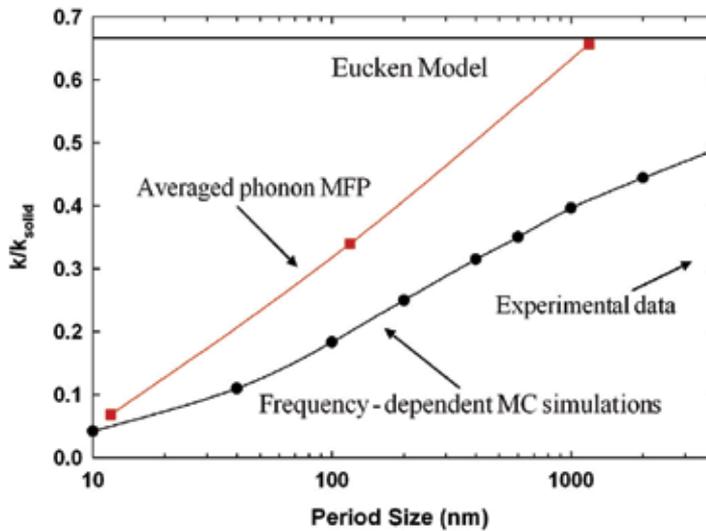


Fig. 6. Normalized in-plane lattice thermal conductivities of porous silicon films as a function of period size (Hao et al., 2009). The circles are the frequency-dependent MC simulation results. The squares are results by solving the BTE for the same 2D unit cell (Miyazaki et al., 2006), with an averaged phonon MFP and isothermal wall boundary condition applied. The triangle is Song's experimental result of a 4.49  $\mu\text{m}$  thick silicon film with similar doping level, 2.3  $\mu\text{m}$  pore diameter, 4  $\mu\text{m}$  pore spacing, and a corresponding porosity  $\Phi = 0.26$  (Song & Chen, 2004). The Eucken model prediction (solid line) is plotted for comparison.

Song and Chen reported a 44.5 W/m·K (also shown in Fig. 6) lattice thermal conductivity at 300 K for a 4.49  $\mu\text{m}$  thick silicon film with similar doping level, 2.3  $\mu\text{m}$  pore diameter, 4  $\mu\text{m}$  pore spacing, and a corresponding porosity  $\Phi = 0.26$  (Song & Chen, 2004). Our current prediction is still higher than this result, which can be mainly attributed to two factors. First, our simulation is 2D while the phonon scattering at the film top and bottom surfaces are mostly diffuse (Mazumder & Majumdar, 2001). We expect such diffuse scattering should further reduce the thermal conductivity. Secondly, during the process of drilling through-film pores, damaged surface layer may be created by the ion bombardment of the deep reactive-ion etching (Song & Chen, 2004), and effectively increase  $A_x$ . The pore shape difference is not expected to be very important because of the phonon diffuse scattering on the pore boundaries and similar porosities. Despite the difference, our simulation shows

that indeed size effects exist even in micron range, consistent with experimental observations. Contribution of long-MFP phonons to the thermal conductivity is significantly suppressed by the scattering of periodically arranged pores, and such influence can be predicted only if frequency-dependent phonon MFPs are considered.

#### 5.4 Silicon nanocomposites

To simulate the phonon transport inside nanocomposites, another critical parameter is the phonon transmissivity across interfaces. The diffuse mismatch model (Swartz, 1989; Chen, 2005) assumes that phonons emitted from an interface cannot tell which side they come from, i.e.,  $1 - T_{12}(\omega) = T_{21}(\omega)$ . Here  $T_{ij}$  denotes the transmissivity from side  $i$  to  $j$ . For an interface with the same material on both sides, we have  $T_{12}(\omega) = T_{21}(\omega)$  by symmetry and thus  $T_{12}(\omega) \equiv 0.5$ . At 575 K, an average phonon transmissivity  $\langle T \rangle = 0.57$  is given by molecular dynamics simulations for silicon grain interfaces (Maiti et al., 1997), which is close to our estimation here.

The studied n-type silicon nanocomposites were synthesized in two steps as described in the literature (Poudel et al., 2008; Ma et al., 2008; Joshi et al., 2008; Wang et al., 2008; Yang et al., 2009; Zhu et al., 2009; Bux et al., 2009). We first prepared nanopowders by high energy ball milling bulk silicon together with the doping element chunks (3% phosphorous in mole). Secondly, we hot-pressed the obtained  $\text{Si}_{1.00}\text{P}_{0.03}$  nanopowders into a bulk disc and measured its TE properties with commercial setups. The ZT of the investigated sample is around 0.55 at 1173 K. The averaged grain size in the sample is determined as 200 nm by transmission electron microscopy. In MC simulations, parameters used for silicon nanocomposites are listed in Table 1. The lattice thermal conductivity  $k_L$  is calculated by subtracting the electronic contribution  $k_e$  from the measured thermal conductivity (Fig. 7). Here  $k_e$  is calculated by the Wiedemann-Franz law,  $k_e = L\sigma T$ , where  $L$  is the computed Lorenz number (Kołodziejczak, 1961),  $\sigma$  is electrical conductivity, and  $T$  is absolute temperature. Due to large grain sizes, we do not consider the grain interface influence on Lorenz numbers. In heavily doped n-type silicon ( $n > 1.0 \times 10^{18} \text{ cm}^{-3}$ ), shallow impurity levels within the band gap start to merge with the conduction band so that the dopants are always completely ionized (Fistul, 1969). With the large band gap of silicon, thermally excited charge carriers contribute negligibly in the investigated temperature range. Therefore, we assume  $n$  is fixed at the dopant concentration for all temperatures. The Fermi level  $E_F$  in Eq. (25) is calculated according to the  $n$  value.

$E_d$ (eV) <sup>a</sup>	$m^*$ ( $m_0$ )	$v_g$ (m/s)	Band gap $E_g$ (eV) <sup>b</sup>	$d$ (kg/m <sup>3</sup> )	$n$ (cm <sup>-3</sup> )
9.5	1.06	6127	$1.17 - 4.73 \times 10^{-4} T^2 / (T + 636)$	2327	$3.93 \times 10^{20}$

<sup>a</sup> Lundstrom, 2000.

<sup>b</sup> Thourmond, 1975.

Table 1. Parameters used for silicon nanocomposites. Temperature-dependent band gap  $E_g$  is used to calculate  $E_F$  based on fixed electron concentration  $n$ . The electron density of states effective mass is calculated for the lowest conduction band valley. The sound velocity is

averaged over TA and LA branches as  $v_g = \left\{ \left[ 2v_{g,TA}^{-1} + v_{g,LA}^{-1} \right] / 3 \right\}^{-1}$ .

Despite the variation of grain shapes inside nanocomposites, we can approximate the structure as packed equal-sized cubes. To avoid conflicts between our boundary condition

and grain interface scatterings of phonons, the chosen computational domain boundary must be away from grain interfaces. Here the computational domain is chosen as a cube consisting of eight adjacent cubes, each of which is  $1/8$  of a cubic grain. In heavily doped polycrystalline silicon, unintentional impurities and imperfections will contribute to impurity scattering and significantly increase the effective  $A$  value. Therefore, we cannot predict  $A$  based on the doping concentration and  $A$  is normally treated as a fitting parameter in analysis. By matching simulation results (empty circles in Fig. 7) with experimental data (filled triangles) at 300 and 573 K, the phonon-impurity scattering coefficient is determined as  $A=1.0\times 10^{-43} \text{ s}^3$ , which is on the same order as previous studies for polycrystalline silicon with similar grain sizes (Uma et al., 2001). The divergence between simulation and experimental results is within 6% at both temperatures. Because of the strong internal scatterings of phonons, we find that grain interface scatterings only slightly affect the phonon transport in this nanocomposite. To compare, we use Eq. (26) to calculate the lattice thermal conductivities of bulk silicon with the same internal scatterings of phonons (dashed line in Fig. 7). At 300 and 573 K, adding grain interfaces (open circles) will only reduce the lattice thermal conductivity by less than 6% from its bulk counterpart (dashed line). The weak influence of grain interfaces can be understood from Fig. 8, which shows the accumulative contributions of phonons with different MFPs to the lattice thermal conductivity. Figure 8 indicates that phonons with MFPs longer than 60 nm contribute negligibly to the lattice thermal conductivity at 300 K. At elevated temperatures, internal scatterings of phonons inside grains are significantly enhanced and thus the influence of grain interfaces becomes even weaker. Without conducting time-consuming MC simulations, we can obtain reasonable agreement between our experimental results and calculated lattice thermal conductivities of heavily doped bulk silicon (Fig. 7).

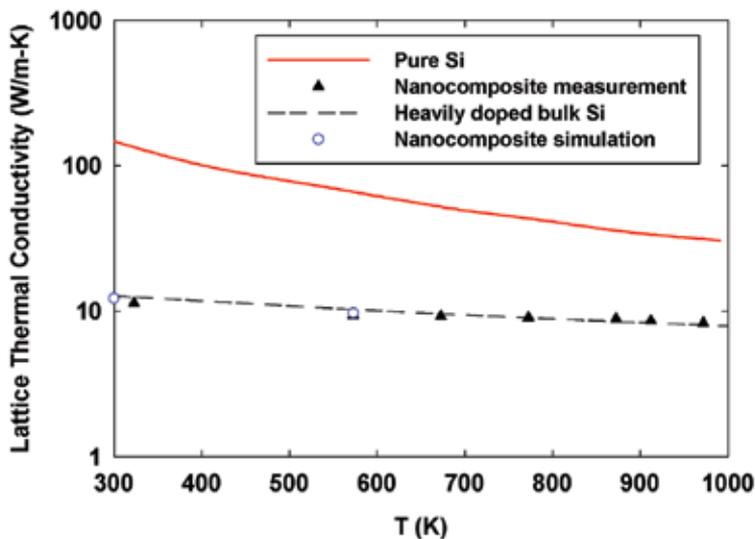


Fig. 7. Lattice thermal conductivities of pure silicon (solid line), measured silicon nanocomposite (filled triangles), calculated heavily doped bulk silicon with electron concentration fixed at  $n=3.93\times 10^{20} \text{ cm}^{-3}$ , and phonon-impurity scattering coefficient  $A=1.0\times 10^{-43} \text{ s}^3$  (dashed line), and simulated 200-nm-grain-size nanocomposite with the same  $n$ ,  $A$  values (empty circles).

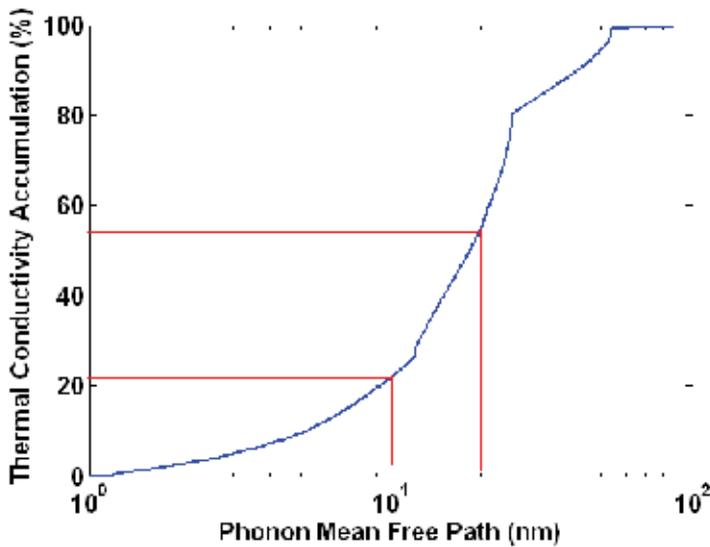


Fig. 8. Room-temperature accumulated thermal conductivity of heavily doped bulk silicon, with electron concentration fixed at  $n=3.93 \times 10^{20} \text{ cm}^{-3}$ , and phonon-impurity scattering coefficient  $A=1.0 \times 10^{-43} \text{ s}^3$ . The accumulation percentage  $a(\Lambda)$  is defined as the percentage of lattice thermal conductivity that is contributed by phonons with MFP less than  $\Lambda$  (Dames & Chen, 2005; Henry & Chen, 2008). From the curve, we can observe that more than 40% of the lattice thermal conductivity is contributed by phonons with MFPs larger than 20 nm, while close to 80% of the lattice thermal conductivity is from phonons with MFPs larger than 10 nm.

To take advantages of the interface scattering of phonons, much smaller grain sizes are required. At 300 K, we find that the room-temperature lattice thermal conductivity can be dropped to 3.0 W/m·K if grain sizes can be reduced to 10 nm, as shown in Fig. 9a. More accurate calculations may further consider the change of  $A$  in such small grains with more complicated strain patterns. Due to dominant grain interface scattering, the lattice thermal conductivities for 10 nm grain sizes are insensitive to temperature variations (Fig. 9b). Because Henry and Chen only provided phonon-phonon scattering rates from 300 to 1000 K for our simulations, the lattice thermal conductivity at 1173 K is linearly extrapolated from simulation results for 973 and 773 K. Slight inaccuracy is expected in this treatment due to the weak temperature dependence of lattice thermal conductivities in 10-nm-grain-size samples.

For charge carrier transport inside nanocomposites, we have developed a model based on the BTE under the relaxation-time approximation (Minnich et al., 2009b). The total relaxation time  $\tau$  for all the scattering mechanisms is obtained by adding up the scattering rates  $\tau_i$  using Matthiessen's rule  $\tau^{-1} = \sum \tau_i^{-1}$  (Lundstrom, 2000; Chen, 2005). Inside a grain, carriers are scattered by acoustic lattice vibrations and ionized impurities (Lundstrom, 2000). On grain interfaces, charge carriers are scattered by a potential barrier created by charges trapped on the interfaces. The grain interface scattering of charge carriers is included into our model by identifying a scattering potential and calculating the corresponding scattering rate, which is then added to the scattering rates of other two mechanisms (Minnich et al., 2009b). Assuming the parameters in Table 1 are unchanged for

10 nm grain sizes, we use the same model to predict  $k_e$  and power factors  $S^2\sigma$  (Fig. 9c) for two typical grain-interface energy barrier heights  $U_g$  (5, 45 meV). The corresponding total thermal conductivities are plotted in Fig. 9b. The power factor and ZT of the reported nano-bulk Si sample (Bux et al., 2009) are also plotted for comparison in Figs. 9c, d. However, its thermal conductivity (not shown in Fig. 9b) cannot be directly compared with the simulated 10-nm-grain-size nanocomposite because this reported sample has a wide distribution of grain sizes (from 10 nm to a few micrometers). In the calculated 10-nm-grain-size nanocomposites, two different barrier heights yield almost identical ZT curves reaching  $ZT \sim 1.02$  at 1173 K (Fig. 9d), which is comparable to conventional SiGe alloys. Figure 9d also shows the predicted ZT of the bulk counterpart, with  $k_L$  computed by Eq. (26) (dashed line in Fig. 7) and electrical properties predicted for  $U_g = 0$  eV in our electron transport model. Its  $ZT \sim 0.56$  at 1173 K is similar to our measured nanocomposites with 200 nm grain sizes.

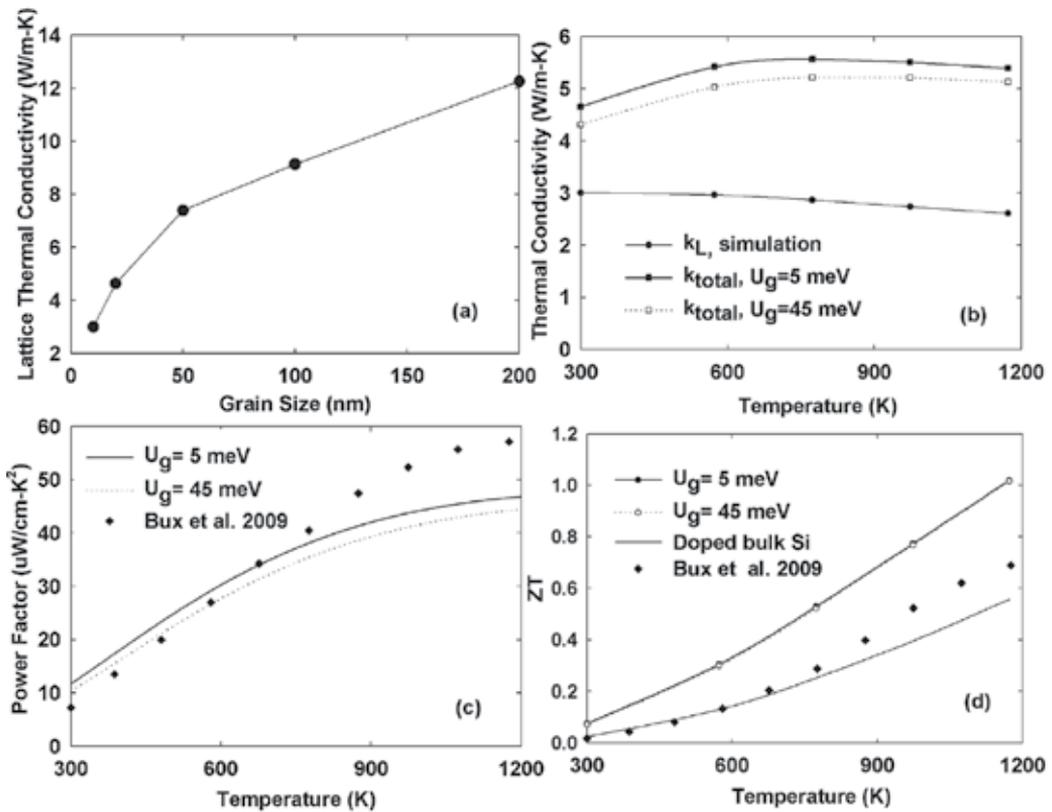


Fig. 9. (a) Grain size dependence of room-temperature  $k_L$  for silicon nanocomposites. (b-d) Temperature-dependent properties of the n-type nano-bulk Si sample (Bux et al., 2009) with grain sizes mainly in the 10-100 nm range, and a 10-nm-grain-size silicon nanocomposite with a grain interface barrier height  $U_g = 5, 45$  meV: (b) thermal conductivities, (c) power factors, (d) thermoelectric figure of merit, compared to that for the heavily doped bulk counterpart with  $k_L$  calculated by Eq. (26) and electrical properties predicted for  $U_g = 0$  eV.

Our studies here show that a ZT above 1.0 is achievable in silicon nanocomposites when the grain size is reduced to 10 nm. Compared with SiGe alloys, silicon nanocomposites eliminate the usage of expensive germanium, which makes it more attractive for commercialization. The main challenge here is to effectively prevent the nanograin growth during hot press and conserve the nano-features to scatter phonons (Poudel et al., 2008; Ma et al., 2008; Wang et al., 2008; Joshi et al., 2008; Yang et al., 2009; Zhu et al., 2009; Bux et al., 2009).

## 6. Summary

In this chapter, frequency-dependent MC simulations are carried out to study the phonon transport inside 2D periodic porous silicon and 3D silicon nanocomposites. A new boundary condition consisting of periodic heat flux with a constant virtual wall temperature is developed for arbitrary periodic structures, enabling accurate thermal conductivity prediction with a single period as the computational domain. This work sets up a framework for future studies of phonon transport in different nanostructures. With accurate information on phonon MFPs in other abundant material systems with good electrical properties, frequency-dependent MC simulations of phonon transport can also be conducted for their corresponding TE nanocomposites. Cheap nanostructured bulk materials with high TE performance can be developed along the way. More broadly, frequency-dependent phonon MC simulations can also provide accurate thermal conductivity predictions for nanostructured materials used in other applications with thermal concerns. Such applications include nanoporous materials for thermal insulation (Chung & Kaviani, 2000), nanoporous electrode materials for batteries (Yamada et al., 2004a; Yamada et al., 2004b; Moriguchi et al., 2006), nanoporous (Logar & Kaučič, 2006) and nanocrystalline (Jurczyk, 2006) hydrogen storage materials.

## 7. Acknowledgments

This work is supported by "Solid State Solar-Thermal Energy Conversion Center" (S<sup>3</sup>TEC), an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Award Number: DE-SC0001299 (G.C.).

## 8. References

- Alexander, M. N. and Holcomb, D. F. (1968). "Semiconductor-to-Metal Transition in n-Type Group IV Semiconductors." *Reviews of Modern Physics*, Vol. 40(4), pp. 815-829.
- Asheghi, M., Kurabayashi, K., Kasnavi, R. and Goodson, K. E. (2002). "Thermal Conduction in Doped Single-Crystal Silicon Films." *Journal of Applied Physics*, Vol. 91(8), pp. 5079-5088.
- Broido, D. A., Malorny, M., Birner, G., Mingo, N. and Stewart, D. A. (2007). "Intrinsic Lattice Thermal Conductivity of Semiconductors from First Principles." *Applied Physics Letters*, Vol. 91(23), pp. 231922-3.
- Bux, S. K., Blair, R. G., Gogna, P. K., Lee, H., Chen, G., Dresselhaus, M. S., Kaner, R. B. and Fleurial, J.-P. (2009). "Nanostructured Bulk Silicon as an Effective Thermoelectric Material." *Advanced Functional Materials*, Vol. 19(15), pp. 2445-2452.

- Cahill, D. G., Ford, W. K., Goodson, K. E., Mahan, G. D., Majumdar, A., Maris, H. J., Merlin, R. and Phillpot, S. R. (2003). "Nanoscale Thermal Transport." *Journal of Applied Physics*, Vol. 93(2), pp. 793-818.
- Callaway, J. (1959). "Model for Lattice Thermal Conductivity at Low Temperatures." *Physical Review*, Vol. 113(4), pp. 1046-1051.
- Chen, G. (2005). *Nanoscale Energy Transport and Conversion: A Parallel Treatment of Electrons, Molecules, Phonons, and Photons*, Oxford University Press, ISBN-13: 978-0195159424, New York, NY, USA.
- Chen, Y., Li, D., Lukes, J. R. and Majumdar, A. (2005). "Monte Carlo Simulation of Silicon Nanowire Thermal Conductivity." *Journal of Heat Transfer*, Vol. 127(10), pp. 1129-1137.
- Chung, J. D. and Kaviani, M. (2000). "Effects of Phonon Pore Scattering and Pore Randomness on Effective Conductivity of Porous Silicon." *International Journal of Heat and Mass Transfer*, Vol. 43(4), pp. 521-538.
- Dames, C. and Chen, G. (2005). "Thermal Conductivity of Nanostructured Thermoelectric Materials." In: *Thermoelectrics Handbook: Macro to Nano*, Rowe, D. M. (Ed.), pp. 42:1-16, CRC Press, ISBN-13: 978-0849322648, Boca Raton, FL, USA.
- Eucken, A. (1932 & 1933). "Thermal Conductivity of Ceramic Refractory Materials: Calculation from Thermal Conductivity of Constituents." *Ceramic Abstracts*, Vol. 11, p. 576, Vol. 12, p. 231.
- Fistul, V. I. (1969). *Heavily Doped Semiconductors*, Plenum Press, ISBN-13: 978-0306303524, New York, NY, USA.
- Garimella, S. V., Fleischer, A. S., Murthy, J. Y., Keshavarzi, A., Prasher, R., Patel, C., Bhavnani, S. H., Venkatasubramanian, R., Mahajan, R., Joshi, Y., Sammakia, B., Myers, B. A., Chorosinski, L., Baelmans, M., Sathyamurthy, P. and Raad, P. E. (2008). "Thermal Challenges in Next-Generation Electronic Systems." *Components and Packaging Technologies, IEEE Transactions on*, Vol. 31(4), pp. 801-815.
- Goldsmid, H. J. (1964). *Thermoelectronic Refrigeration*, Plenum Press, ISBN-13: 978-0306301780, London, UK.
- Goodson, K. E., Ju, Y. S. and Asheghi, M. (1997). "Thermal Phenomena in Semiconductor Devices and Interconnects." In: *Microscale Energy Transport*, Tien, C. L., Majumdar, A. and Gerner, F. M. (Eds.), pp. 229-293, Taylor & Francis, ISBN-13: 978-1560324591, Washington, DC.
- Hao, Q., Chen, G. and Jeng, M.-S. (2009). "Frequency-Dependent Monte Carlo Simulations of Phonon Transport in Two-Dimensional Porous Silicon with Aligned Pores." *Journal of Applied Physics*, Vol. 106(11), pp. 114321/1-10.
- Hao, Q., Zhu, G., Joshi, G., Wang, X., Minnich, A., Ren, Z. and Chen, G. (2010). "Theoretical Studies on the Thermoelectric Figure of Merit of Nano-Grained Bulk Silicon." *Applied Physics Letters*, Vol. 97(6), pp. 063109/1-3.
- Henry, A. and Chen, G. (2008). "Spectral Phonon Transport Properties of Silicon Based on Molecular Dynamics Simulations and Lattice Dynamics." *Journal of Computational and Theoretical Nanoscience*, Vol. 5, pp. 141-152.
- Hochbaum, A. I., Chen, R., Delgado, R. D., Liang, W., Garnett, E. C., Najarian, M., Majumdar, A. and Yang, P. (2008). "Enhanced Thermoelectric Performance of Rough Silicon Nanowires." *Nature*, Vol. 451(7175), pp. 163-167.

- Holland, M. G. (1963). "Analysis of Lattice Thermal Conductivity." *Physical Review*, Vol. 132(6), pp. 2461-2471.
- Hsu, K. F., Loo, S., Guo, F., Chen, W., Dyck, J. S., Uher, C., Hogan, T., Polychroniadis, E. K. and Kanatzidis, M. G. (2004). "Cubic  $\text{AgPb}_m\text{SbTe}_{2+m}$ : Bulk Thermoelectric Materials with High Figure of Merit." *Science*, Vol. 303(5659), pp. 818-821.
- Jeng, M.-S., Yang, R., Song, D. and Chen, G. (2008). "Modeling the Thermal Conductivity and Phonon Transport in Nanoparticle Composites Using Monte Carlo Simulation." *Journal of Heat Transfer*, Vol. 130(4), pp. 042410-11.
- Joshi, G., Lee, H., Lan, Y., Wang, X., Zhu, G., Wang, D., Gould, R. W., Cuff, D. C., Tang, M. Y., Dresselhaus, M. S., Chen, G. and Ren, Z. (2008). "Enhanced Thermoelectric Figure-of-Merit in Nanostructured p-Type Silicon Germanium Bulk Alloys." *Nano Letters*, Vol. 8(12), pp. 4670-4674.
- Jurczyk, M. (2006). "Nanocrystalline Materials for Hydrogen Storage." *Journal of Optoelectronics and Advanced Materials*, Vol. 8, pp. 418-424.
- Kane, E. O. (1957). "Band Structure of Indium Antimonide." *Journal of Physics and Chemistry of Solids*, Vol. 1(4), pp. 249-261.
- Klemens, P. G. (1955). "The Scattering of Low-Frequency Lattice Waves by Static Imperfections." *Proceedings of the Physical Society. Section A*, Vol. 68(12), pp. 1113-1128.
- Klitsner, T., VanCleve, J. E., Fischer, H. E. and Pohl, R. O. (1988). "Phonon Radiative Heat Transfer and Surface Scattering." *Physical Review B*, Vol. 38(11), pp. 7576-7594.
- Kolodziejczak, J. (1961). "Transport of Current Carriers in n-Type Indium Antimonide at Low Temperatures." *Acta Physica Polonica*, Vol. 20, pp. 289-302.
- Lacroix, D., Joulain, K. and Lemonnier, D. (2005). "Monte Carlo Transient Phonon Transport in Silicon and Germanium at Nanoscales." *Physical Review B*, Vol. 72(6), pp. 064305/1-11.
- Lacroix, D., Joulain, K., Terris, D. and Lemonnier, D. (2006). "Monte Carlo Simulation of Phonon Confinement in Silicon Nanostructures: Application to the Determination of the Thermal Conductivity of Silicon Nanowires." *Applied Physics Letters*, Vol. 89(10), pp. 103104-3.
- Lee, J. H., Grossman, J. C., Reed, J. and Galli, G. (2007). "Lattice Thermal Conductivity of Nanoporous Si: Molecular Dynamics Study." *Applied Physics Letters*, Vol. 91(22), pp. 223110-3.
- Lindsay, L., and Broido, D.A. (2008). "Three-Phonon Phase Space and Lattice Thermal Conductivity in Semiconductors." *Journal of Physics: Condensed Matter*, Vol. 20, pp. 165209/1-6.
- Logar, N. Z. and Kaučič, V. (2006). "Nanoporous Materials: From Catalysis and Hydrogen Storage to Wastewater Treatment." *Acta Chimica Slovenica*, Vol. 53, pp. 117-135.
- Lundstrom, M. (2000). *Fundamentals of Carrier Transport*, Oxford University Press, ISBN-13: 978-0521631341, Cambridge, UK.
- Ma, Y., Hao, Q., Poudel, B., Lan, Y., Yu, B., Wang, D., Chen, G. and Ren, Z. (2008). "Enhanced Thermoelectric Figure-of-Merit in p-Type Nanostructured Bismuth Antimony Tellurium Alloys Made from Elemental Chunks." *Nano Letters*, Vol. 8(8), pp. 2580-2584.

- Maiti, A., Mahan, G. D. and Pantelides, S. T. (1997). "Dynamical Simulations of Nonequilibrium Processes - Heat Flow and the Kapitza Resistance across Grain Boundaries." *Solid State Communications*, Vol. 102(7), pp. 517-521.
- Mazumder, S. and Majumdar, A. (2001). "Monte Carlo Study of Phonon Transport in Solid Thin Films Including Dispersion and Polarization." *Journal of Heat Transfer*, Vol. 123(4), pp. 749-759.
- McConnell, A. D., and Goodson, K.E. (2005). "Thermal Conduction in Silicon Micro and Nanostructures." *Annual Review of Heat Transfer*, Vol. 14, pp. 129-168.
- McGaughey, A. J. H. and Kaviani, M. (2005). "Observation and Description of Phonon Interactions in Molecular Dynamics Simulations." *Physical Review B*, Vol. 71(18), pp. 184305/1-11.
- Minnich, A. J., Dresselhaus, M. S., Ren, Z. F. and Chen, G. (2009a). "Bulk Nanostructured Thermoelectric Materials: Current Research and Future Prospects." *Energy & Environmental Science*, Vol. 2(5), pp. 466-479.
- Minnich, A. J., Lee, H., Wang, X. W., Joshi, G., Dresselhaus, M. S., Ren, Z. F., Chen, G. and Vashaee, D. (2009b). "Modeling Study of Thermoelectric SiGe Nanocomposites." *Physical Review B*, Vol. 80(15), pp. 155327/1-11.
- Miyazaki, K., Arashi, T., Makino, D. and Tsukamoto, H. (2006). "Heat Conduction in Microstructured Materials." *Components and Packaging Technologies, IEEE Transactions on*, Vol. 29(2), pp. 247-253.
- Moriguchi, I., Hidaka, R., Yamada, H., Kudo, T., Murakami, H. and Nakashima, N. (2006). "A Mesoporous Nanocomposite of TiO<sub>2</sub> and Carbon Nanotubes as a High-Rate Li-Intercalation Electrode Material." *Advanced Materials*, Vol. 18(1), pp. 69-73.
- Narumanchi, S. V. J., Murthy, J. Y. and Amon, C. H. (2006). "Boltzmann Transport Equation-Based Thermal Modeling Approaches for Hotspots in Microelectronics." *Heat and Mass Transfer*, Vol. 42(6), pp. 478-491.
- Pattamatta, A. and Madnia, C. K. (2009). "Modeling Heat Transfer in Bi<sub>2</sub>Te<sub>3</sub>-Sb<sub>2</sub>Te<sub>3</sub> Nanostructures." *International Journal of Heat and Mass Transfer*, Vol. 52(3-4), pp. 860-869.
- Peterson, R. B. (1994). "Direct Simulation of Phonon-Mediated Heat Transfer in a Debye Crystal." *Journal of Heat Transfer*, Vol. 116(4), pp. 815-822.
- Poudel, B., Hao, Q., Ma, Y., Lan, Y., Minnich, A., Yu, B., Yan, X., Wang, D., Muto, A., Vashaee, D., Chen, X., Liu, J., Dresselhaus, M. S., Chen, G. and Ren, Z. (2008). "High-Thermoelectric Performance of Nanostructured Bismuth Antimony Telluride Bulk Alloys." *Science*, Vol. 320(5876), pp. 634-638.
- Prasher, R. (2006). "Thermal Conductivity of Composites of Aligned Nanoscale and Microscale Wires and Pores." *Journal of Applied Physics*, Vol. 100(3), pp. 034307-9.
- Randrianalisoa, J. and Baillis, D. (2008). "Monte Carlo Simulation of Steady-State Microscale Phonon Heat Transport." *Journal of Heat Transfer*, Vol. 130(7), pp. 072404-13.
- Song, D. and Chen, G. (2004). "Thermal Conductivity of Periodic Microporous Silicon Films." *Applied Physics Letters*, Vol. 84(5), pp. 687-689.
- Swartz, E. T. and Pohl, R. O. (1989). "Thermal Boundary Resistance." *Reviews of Modern Physics*, Vol. 61(3), pp. 605-668.

- Tang, J., Wang, H.-T., Lee, D. H., Fardy, M., Huo, Z., Russell, T. P. and Yang, P. (2010). "Holey Silicon as an Efficient Thermoelectric Material." *Nano Letters*, Vol. 10(10), pp. 4279-4283.
- Thurmond, C. D. (1975). "The Standard Thermodynamic Functions for the Formation of Electrons and Holes in Ge, Si, GaAs, and Gap." *Journal of The Electrochemical Society*, Vol. 122(8), pp. 1133-1141.
- Tian, W. and Yang, R. (2007). "Thermal Conductivity Modeling of Compacted Nanowire Composites." *Journal of Applied Physics*, Vol. 101(5), pp. 054320-5.
- Turney, J. E., Landry, E. S., McGaughy, A. J. H. and Amon, C. H. (2009). "Predicting Phonon Properties and Thermal Conductivity from Anharmonic Lattice Dynamics Calculations and Molecular Dynamics Simulations." *Physical Review B*, Vol. 79(6), pp. 064301/1-12.
- Uma, S., McConnell, A. D., Asheghi, M., Kurabayashi, K. and Goodson, K. E. (2001). "Temperature Dependent Thermal Conductivity of Undoped Polycrystalline Silicon Layers." *International Journal of Thermophysics*, Vol. 22, pp. 605-616.
- Wang, X. W., Lee, H., Lan, Y. C., Zhu, G. H., Joshi, G., Wang, D. Z., Yang, J., Muto, A. J., Tang, M. Y., Klatsky, J., Song, S., Dresselhaus, M. S., Chen, G. and Ren, Z. F. (2008). "Enhanced Thermoelectric Figure of Merit in Nanostructured n-Type Silicon Germanium Bulk Alloy." *Applied Physics Letters*, Vol. 93(19), pp. 193121-3.
- Yamada, H., Yamato, T., Moriguchi, I. and Kudo, T. (2004a). "Porous TiO<sub>2</sub> (Anatase) Electrodes for High-Power Batteries." *Chemistry Letters*, Vol. 33(12), pp. 1548-1549.
- Yamada, H., Yamato, T., Moriguchi, I. and Kudo, T. (2004b). "Interconnected Macroporous TiO<sub>2</sub> (Anatase) as a Lithium Insertion Electrode Material." *Solid State Ionics*, Vol. 175, pp. 195-198.
- Yamamoto, A., Takazawa, H. and Ohta, T. (1999). "Thermoelectric Transport Properties of Porous Silicon Nanostructure." *Proceedings of 18th International Conference on Thermoelectrics*, pp. 428-431, ISBN 0-7803-5451-6, Baltimore, MD, USA, Aug 1999 - Sep 1999, IEEE Publications Office, Los Alamitos, CA, USA.
- Yamanouchi, C., Mizuguchi, K. and Sasaki, W. (1967). "Electric Conduction in Phosphorus Doped Silicon at Low Temperatures." *Journal of the Physical Society of Japan*, Vol. 22(3), pp. 859-864.
- Yang, J., Hao, Q., Wang, H., Lan, Y. C., He, Q. Y., Minnich, A., Wang, D. Z., Harriman, J. A., Varki, V. M., Dresselhaus, M. S., Chen, G. and Ren, Z. F. (2009). "Solubility Study of Yb in n-Type Skutterudites Yb<sub>x</sub>Co<sub>4</sub>Sb<sub>12</sub> and Their Enhanced Thermoelectric Properties." *Physical Review B*, Vol. 80(11), pp. 115329.
- Yang, R. and Chen, G. (2004). "Thermal Conductivity Modeling of Periodic Two-Dimensional Nanocomposites." *Physical Review B*, Vol. 69(19), pp. 195316/1-10.
- Yang, R., Chen, G. and Dresselhaus, M. S. (2005). "Thermal Conductivity of Simple and Tubular Nanowire Composites in the Longitudinal Direction." *Physical Review B*, Vol. 72(12), pp. 125418/1-7.
- Yu, J.-K., Mitrovic, S., Tham, D., Varghese, J. and Heath, J. R. (2010). "Reduction of Thermal Conductivity in Phononic Nanomesh Structures." *Nature Nanotechnology*, advance online publication.
- Zhu, G. H., Lee, H., Lan, Y. C., Wang, X. W., Joshi, G., Wang, D. Z., Yang, J., Vashaee, D., Guilbert, H., Pillitteri, A., Dresselhaus, M. S., Chen, G. and Ren, Z. F. (2009).

"Increased Phonon Scattering by Nanograins and Point Defects in Nanostructured Silicon with a Low Concentration of Germanium." *Physical Review Letters*, Vol. 102(19), pp. 196803/1-4.

Ziman, J. M. (1956 & 1957). "The Effect of Free Electrons on Lattice Conduction." *Philosophical Magazine*, Vol. 1(2), pp. 191-198, Vol. 2(14), pp. 292.

# Performance Analysis of Adaptive GPS Signal Detection in Urban Interference Environment using the Monte Carlo Approach

V. Behar<sup>1</sup>, Ch. Kabakchiev<sup>2</sup>, I. Garvanov<sup>3</sup> and H. Rohling<sup>4</sup>

<sup>1</sup>*Institute of Information and Communication Technologies-BAS,*

<sup>2</sup>*Sofia University "Sv. Kl Ohridski",*

<sup>3</sup>*State University of Library Studies and Information Technologies,*

<sup>4</sup>*Technical University Hamburg-Harburg,*

<sup>1,2,3</sup>*Bulgaria*

<sup>4</sup>*Germany*

## 1. Introduction

The Global Positioning System (GPS) has been designed to provide precision location estimates for various military and civil applications. Each of satellites transmits digitally coded data, and GPS receivers demodulate these signals from four or more satellites simultaneously in order to generate three time-difference-of-arrival estimates, allowing the user to measure the range to three satellites, and, as a result, to determine his position. Since a direct sequence spread spectrum (DSSS) signal is used in transmission, relatively low powers can be transmitted by the satellites and still to have adequate signal-to-noise ratio (SNR) for accurate position estimation. In fact, these signals have SNR of between -15dB and -30dB. In civil applications, therefore, the key to achieve the precise position estimation performance is the processing of very weak DSSS signals from satellites that contain coarse acquisition (C/A) digitally coded data. In order to extract the information from a GPS signal, the presence of the coarse acquisition (C/A) code must be detected in the signal arriving at the input of a GPS receiver. In each channel of a GPS receiver, the algorithm for detection of the C/A code that identifies the corresponding satellite is carried out by cross-correlation and thresholding procedures. The objective of the detection algorithm is to search for the presence of the GPS signal over a frequency range that covers every possible expected Doppler frequency. The intensity maximum in the time discrete  $n$  and the frequency bin  $m$  gives the beginning point of the C/A code in 200 ns resolution and the carrier frequency in 1 kHz resolution if the signal maximum is above the predetermined threshold of detection. Once the GPS signal is found, two important parameters, the beginning point of the C/A code period and the carrier frequency of the input signal, are measured. Next this information is passed on to the tracking algorithm in order to extract the navigation data (Tsui, 2005).

Whatever GPS signals have some degree of antijamming protection built into the signal structure itself, the weak signal strength of the received signal makes it easy for strong broadband interference to overcome the antijamming protection of the C/A code signal

(Sklar, 2003). If a strong broadband jamming source is nearby, the receiver noise may rise to the level where the SNR at the correlator output is below the threshold value required for tracking. In that case the capability of the correlator to detect the C/A code is very seriously degraded. Multipath is the other limiting factor in many GPS applications that affects pseudorange and carrier phase estimates. Signal multipath is the phenomenon where a satellite signal arrives at the receiver antenna after being reflected from different surfaces or buildings (Soubielle et al., 2002).

Various approaches can be used to mitigate GPS interference before signal processing in a GPS receiver (Fu et al., 2003; Sklar, 2003). One of them is to use different beamforming techniques for broadband nulling having in mind that the satellite signals and interfering signals usually originate from different spatial locations. The conventional (delay-and-sum) beamforming procedure is performed by the simplest non-adaptive algorithm, in which all weights have equal magnitudes and the phases are selected to steer the array in particular direction (Van Trees, 2002). Such a beamformer has unity response in each look direction, and in conditions of no directional interferences, the beamformer provides maximum SNR but it is not effective in the presence of directional jamming signals, intentional or unintentional. The other beamformers such as a Minimum Variance Distortionless Response (MVDR) beamformer can overcome this problem by suppressing interfering signals from off-axis directions (Vouras & Freburger, 2008; Tummonery, 2005). To suppress jamming signals, the MDVR beamformer does not require a priori information about them but only the information for the direction-of-arrival of a desired signal. The capability of the MVDR beamformer to improve the detectability of radar targets in conditions of strong jamming is investigated in (Behar & Kabakchiev, 2009; Behar et al, 2010). The impact of different factors on the capability of the adaptive MVDR beamformer to mitigate broadband interference at the input of GPS receivers is studied in (Behar et al., 2009; Behar et al., 2010). FPGA implementation of the MVDR QR-based beamformer for broadband interference suppression in satellite navigation receivers is proposed in (Ganchosov, 2009). The performance of each beamforming method can be applied to different antenna array configurations. The antenna elements are put together in a known geometry, which is usually uniform - Uniform Linear Arrays (ULA), Uniform Rectangular Arrays (URA) or Uniform Circular Arrays (UCA) (Ioannides & Balanis, 2005; Moelker, 1996). Two configurations, URA and UCA, with the elements extended in two dimensions enable to control the beam pattern in both azimuth and elevation, and for that reason they can be used for implementation of beamforming in GPS receivers. The smallest inter-element spacing in antenna arrays is usually equal to or slightly less than half a wavelength of the satellite carrier frequency ( $\lambda/2$ ) in order to avoid the problem of "spatial under-sampling". However, an interelement spacing smaller than  $\lambda/2$  increases the risk of mutual coupling between antenna elements.

As a rule, the standard C/A code detection performance is designed having in mind the detection on the background of a receiver noise only (Tsui, 2005). However, the application of modern radar approaches to signal processing in a GPS receiver can overcome the problem associated with detection of weak GPS signals in conditions of strong urban interference. In this chapter we present our original idea to combine three different approaches in a new three-stage algorithm for detection of the C/A code in heavy urban noise environment (GPS MDVR CFAR). The proposed GPS MDVR CFAR detection algorithm includes three processing stages: (i) - adaptive Minimum Variance Distortionless Response (MVDR) beamforming algorithm applied to the software GPS receiver input; (ii) -

circular cross-correlation algorithm performed in the frequency domain; (iii) - CFAR signal thresholding algorithm applied to the cross-correlator output for maintaining the constant probability of false alarm. The design of the GPS MDVR CFAR detection algorithm is based on our experience in the research of algorithms for broadband/ pulse jamming suppression (Behar & Kabakchiev, 2009; Behar et al., 2010, Behar et al., 2009; Ganchosov et al., 2009) and target detection (Behar et al., 2000; Garvanov et al., 2003; Kabakchiev et al., 2010; Behar et al., 2010) in GPS and radar applications. In this chapter, the effectiveness of the GPS MDVR CFAR detection algorithm is expressed in terms of three quality parameters: the signal-to-noise-plus-interference ratio (SINR) improvement factor statistically estimated at the beamformer output; the post correlation signal-to-noise ratio (SNR) statistically estimated at the cross-correlator output and, finally, the probability of detection estimated at the CFAR detector output.

The three quality parameters of the GPS MDVR CFAR detection algorithm are evaluated using the Monte Carlo approach. The statistical estimates of the corresponding quality factors are evaluated for each stage of the detection algorithm. The objectives of the Monte Carlo analysis are: (i)- to analyze the capability of two beamformers, non-adaptive (conventional) and adaptive (MVDR), to mitigate broadband radio frequency interference (RFI) at the navigation receiver input and as a result to improve the post correlation SNR and, thus, increasing the probability of detection of the C/A code in conditions of strong jamming; (ii) - to evaluate the influence of several important factors on the performance of the joint three-stage detection algorithm. These factors include: interference intensity expressed in terms of the interference-to-signal ratio (ISR); planar array configuration (*URA* and *UCA*); number of array elements; sampling rate of the incoming data; angular errors in satellite location known as steering vector mismatch; reference window length used in a CFAR detector.

## 2. Antenna array geometry

Antenna arrays are composed of many antenna elements in order to create a unique radiation pattern in the desired direction. The antenna elements are put together in a known geometrical structure with uniform interelement spacing - Uniform Linear Arrays (ULA), Uniform Rectangular Arrays (URA) or Uniform Circular Arrays (UCA) (Van Trees, 2002; Ioannidis & Balanis, 2005). Since a ULA beam pattern can be controlled in only one dimension (azimuth), in GPS applications, only URA or UCA configurations with the elements extended in two dimensions should be used in order to control the beam pattern in two angular dimensions (azimuth and elevation).

### 2.1 URA configuration

In URA antenna arrays, all elements are extended in the  $x$ - $y$  plane. There are  $M_X$  elements in the  $x$ -direction and  $M_Y$  elements in the  $y$ -direction creating an array of  $(M_X \times M_Y)$  elements. All elements are uniformly spaced  $d$  apart in both directions. Such a rectangular array can be viewed as  $M_Y$  uniform linear arrays of  $M_X$  elements or  $M_X$  uniform linear arrays of  $M_Y$  elements. Usually, the first antenna array element is considered as the origin of Cartesian coordinates in (Fig. 1). The direction of a signal arriving from azimuth  $\varphi$  and elevation  $\theta$  can be described with a unit vector  $e$  in Cartesian coordinates as:

$$e(\varphi, \theta) = (e_x, e_y, e_z) = (\cos\theta\sin\varphi, \cos\theta\cos\varphi, \sin\theta) \quad (1)$$

The element number  $m(i,k)$  of an antenna array is calculated as:

$$m(i,k) = (i-1)M_X + k, \quad i = 1 \div M_Y, \quad k = 1 \div M_X \quad (2)$$

The vector in the direction of element  $m(i,k)$  can be described in Cartesian coordinates as:

$$r_{m(i,k)} = (d(i-1), d(k-1), 0) \quad (3)$$

In (3), indexes  $i$  and  $k$  denote the element position along the  $y$ - and the  $x$ -axis, respectively. If the first element of a rectangular array is a reference element, the path-length difference  $d_{m(i,k)}$  for a signal incident at element  $m(i,k)$  can be defined as a projection of the vector  $r_{m(i,k)}$  on the signal direction vector  $e$ :

$$d_{m(i,k)} = e^T \cdot r_{m(i,k)} = \cos\theta \cdot d \cdot [\sin\varphi(i-1) + \cos\varphi(k-1)] \quad (4)$$

Therefore, the URA response vector  $a_c$  in the direction  $(\varphi, \theta)$  takes the form:

$$a_c(\varphi, \theta) = [1, \exp(j\frac{2\pi}{\lambda}d_2), \dots, \exp(j\frac{2\pi}{\lambda}d_{m(i,k)}), \dots, \exp(j\frac{2\pi}{\lambda}d_M)] \quad (5)$$

where  $M = M_X \times M_Y$ .

## 2.2 UCA configuration

In UCA antenna arrays, all elements are arranged along the ring of radius  $r$  (Fig. 2). The ring contains  $M$  array elements. Since these elements are uniformly spaced along the ring, they have an interelement angular spacing  $\Delta\varphi = 2\pi/M$  and a linear interelement spacing  $d = 2r\pi/M$ . It is usually assumed that the first antenna element is located on the  $y$ -axis, and the ring center is the origin of Cartesian coordinates. The vector in the direction of the  $m$ th array element can be defined in Cartesian coordinates as:

$$r_m = (r \sin\varphi_m, r \cos\varphi_m, 0), \quad \text{where } \varphi_m = 2\pi(m-1)/M \quad (6)$$

The unit vector  $e(\varphi, \theta)$  in the direction of a signal source is given by (1). If the ring center serves as a reference point, the propagation path-length difference  $d_m$  for a signal incident at element  $m$  can be defined as a projection of the vector  $r_m$  on the direction vector  $e$ :

$$d_m = e^T \cdot r_m = d \cdot \cos\theta (\sin\varphi \sin\varphi_m + \cos\varphi \cos\varphi_m) = d \cos\theta \cos(\varphi - \varphi_m) \quad (7)$$

Therefore, the UCA response vector  $a_c$  takes the form:

$$a_c(\varphi, \theta) = [\exp(j\frac{2\pi}{\lambda}d_1), \exp(j\frac{2\pi}{\lambda}d_2), \dots, \exp(j\frac{2\pi}{\lambda}d_m), \dots, \exp(j\frac{2\pi}{\lambda}d_M)] \quad (8)$$

where  $d_m$  is expressed by (7) for  $m=1, 2, \dots, M$ .

The vectors  $a_c$  in (5 and 8) are often called steering vectors that describe the array response to a signal arriving from direction  $(\varphi, \theta)$ .

The special case of a circular antenna array is a 7-element antenna array, where the first element is located in the antenna array centre, and the other six elements are arranged in a circle relative to the first element (UCA-7).

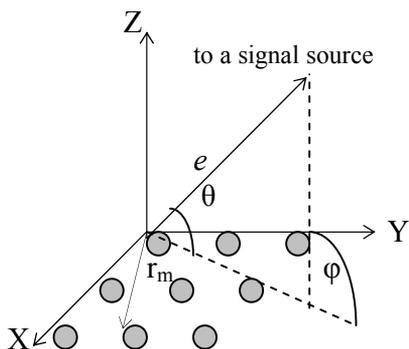


Fig. 1. URA configuration

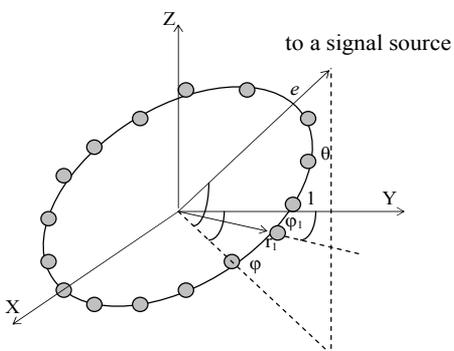


Fig. 2. UCA configuration

### 3. Software-based GPS receiver

The structure of a software-based GPS receiver is shown in Fig. 3. The GPS receiver processes signals received from satellites that are in view, and then uses the extracted information to determine and display the user position, velocity, and time (Tsui, 2005).

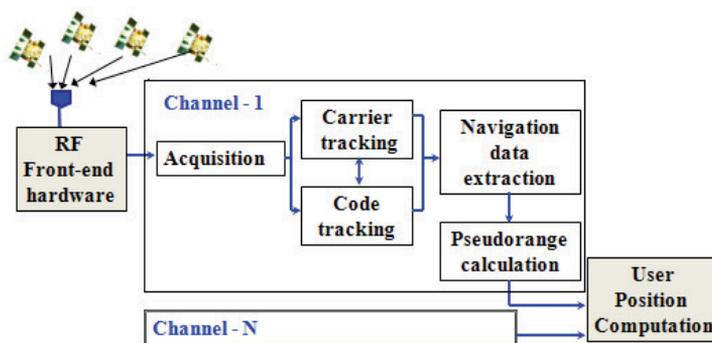


Fig. 3. Architecture of a software-based GPS receiver

The GPS receiver is a multi-channel device, where one channel processes the incoming signal from one satellite. The first two blocks of a GPS receiver are an antenna and a RF

front-end section that includes several devices, which are usually implemented in hardware. In the RF front-end section, the incoming signal is down converted from the RF frequency to an intermediate frequency (IF) in several stages. The down converters reduce the carrier frequency from GHz to a couple MHz. The last stage of the RF front-end section is an Analog-to Digital Converter (ADC), where the IF signals are sampled at a suitable sampling frequency and digitized. In the next blocks of the GPS receiver, the incoming signal is processed in multiple (from 8 to 12) parallel channels, where each channel acquires and dynamically tracks signals from one visible satellite.

As shown in Fig. 3, in each channel the acquisition block detects the signal of a certain satellite, then the tracking block is used to find the phase transition of the navigation data, and the navigation data is obtained from the navigation data phase transition. Further, the ephemeris data and pseudoranges are calculated from the navigation data. The ephemeris data is used to calculate the satellite position and, finally, the user position is obtained from the satellite position and the pseudoranges. Minimum four channels for tracking the signals of four satellites are required to determine three position coordinates and the receiver clock offset.

In a conventional hardware-based receiver, the acquisition and tracking blocks are implemented in an IC chip, and the algorithms implemented in the chip are not available for the user. In a software-based GPS receiver, however, these blocks are implemented in software and hence the user has a free access to these algorithms and can update them and exercise control over them. This is the difference between the software-based GPS receiver and a conventional hardware-based GPS receiver.

#### 4. Signal model

GPS signals are transmitted at two radio frequency bands - L1 (1572.42 MHz) and L2 (1227.6 MHz). Each satellite transmits two unique codes. The first of them is the coarse acquisition code (C/A), and the second code is the encrypted precision code (P(Y)). The signal transmitted at L1 frequency contains the coarse acquisition (C/A) code, precision (P) code and navigation data while the signal transmitted at L2 frequency contains P-code only. We limit our discussion within the signal transmitted at L1 frequency that contains the C/A code only because the signal transmitted at frequency L2 and containing P-code serves for military purposes only and the civilian community does not have a free access to P-code. The C/A code modulated signal is a BPSK modulated signal. The null-to-null frequency bandwidth of the main lobe of this signal spectrum is 2.046 MHz. The total code period contains 1023 chips, and 1023 chips last 1ms with a chip rate of 1.023 MHz. The GPS C/A coded signals belong to the family of Gold codes, and they are unique for every satellite. The main property of the C/A code is that it has the best cross-correlation characteristic. The C/A code modulated signal transmitted by each satellite is the product of three signals: (i)-the carrier signal with frequency L1; (ii)-the navigation data with a bit rate of 50bps; (iii)-the C/A code that is unique for each satellite.

The signal received by an antenna array of a GPS receiver is composed of the satellite signal, thermal noise and a variety of interference. In conditions of jamming, the complex valued samples of the received signal at time instant  $k$  can be mathematically described as:

$$x(k) = a_c s(k) + \sum_{l=1}^L b_l j_l(k) + n(k) \quad (9)$$

where  $x(k)$  is the  $(M \times 1)$ -element signal vector,  $s(k)$  is the received satellite signal sample,  $j_l(k)$  is the  $l$ th broadband jamming sample,  $a_c$  and  $b_l$  are the  $(M \times 1)$ -element antenna array response vectors in the direction of a satellite signal and the  $l$ th broadband interference, respectively,  $n(k)$  is the noise sample and  $L$  is the number of broadband interference sources. The signal received from a satellite can be described as:

$$s(k) = \sqrt{SNR} \cdot c(k) \cdot \cos(2\pi f_0 t_k + \varphi) \cdot d(k) \tag{10}$$

where  $SNR$  is the received signal-to-noise ratio,  $c(k)$  is the C/A code of length  $(20 \times 1023)$ , unique for each satellite,  $d(k)$  is the navigation data bit which remains constant over 20 periods of the C/A code, and  $f_0$  is the carrier intermediate frequency. The jamming signal  $j(k)$  occupies the entire receiver bandwidth and can be modelled as bandlimited additive white Gaussian noise (AWGN) with zero mean:

$$j(k) = \sqrt{INR} \cdot N(0,1) \tag{11}$$

where  $INR$  is the interference-to-noise ratio. The noise sample  $n(k)$  can be modeled as additive white Gaussian noise with zero mean and unity variance. In GPS applications, when the receiver is on the ground, the input SNR value depends on the RF-bandwidth of the receiver front-end and is typically around -20dB (2.046 MHz C/A code bandwidth).

### 5. GPS signal detection

As a rule, the standard C/A code detection performance is designed having in mind the signal detection on the background of a receiver noise only (Tsui, 2005). However, the application of modern radar approaches to signal processing in a GPS receiver can overcome the problems associated with detection of weak GPS signals in conditions of strong urban interference. In this section we present our original idea to combine three different approaches in a new three-stage algorithm for detection of the C/A code in heavy urban noise environment (GPS MDVR CFAR) (Fig.4).

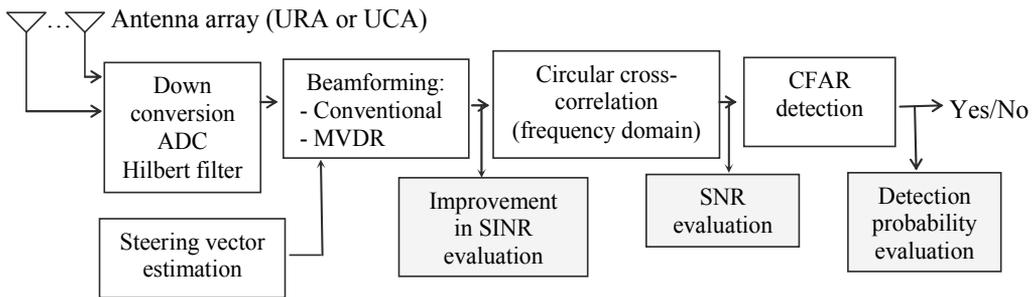


Fig. 4. Flow-chart of the signal processing and the evaluation process

As shown in Fig.4, in the RF front-end section of each antenna element, the GPS signal is down converted from the RF frequency to an intermediate frequency (IF), sampled at a suitable sampling frequency and digitized. After Hilbert filtration the signal is converted into complex form. The complex-valued signal is then performed by the new GPS MDVR CFAR detection algorithm that includes three processing blocks. The first of them is a

beamformer that mitigates broadband interference before detection of the C/A code. The beamformer can be realized as a non-adaptive block, where the conventional “delay-and-sum” method is implemented, or as an adaptive one, where some adaptive method is implemented - for example, the Minimum Variance Distortionless Response (MVDR) method. In the second block the circular cross-correlation procedure is performed in the frequency domain. The incoming signal is correlated with the local signal replica in order to identify the visible satellite in the incoming data and then find the beginning point of the C/A code and estimate the rough Doppler shift. The third block is used to detect the presence of the C/A code at the cross-correlator output while maintaining the constant probability of false alarm. The performance of the modified Cell Averaging CFAR (CA CFAR) detector is considered to be used in this block.

### 5.1 Beamforming methods

The digital beamformer increases the gain in the direction of arrival of the desired signal, and decreases the gain in all other directions (interference). For an antenna array composed of  $M$  elements, the output signal  $Y$  is formed as a weighted sum of signals  $X$  arrived at the antenna input (Fig.5):

$$Y = W^H X, \quad (12)$$

In (12),  $X$  is the input signal matrix of size  $(M \times N)$ ,  $N$  is the number of time samples,  $W$  is the complex valued weight vector of size  $M$ , and  $(.)^H$  denotes conjugate transpose.

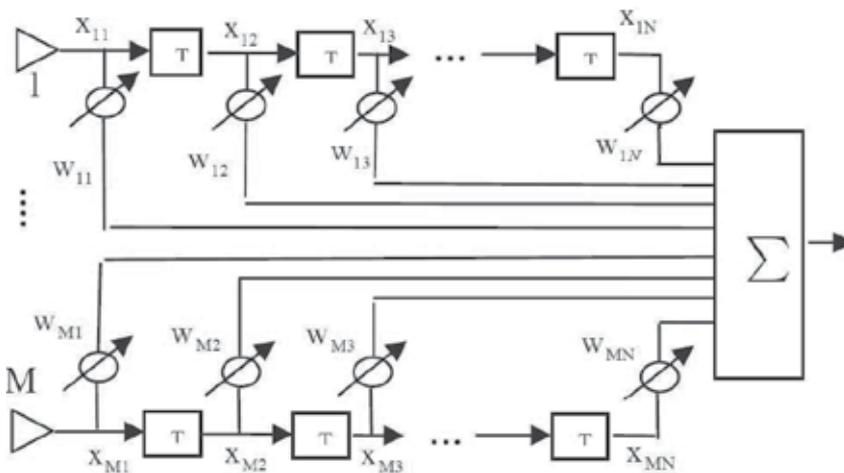


Fig. 5. Block diagram of the beamforming technique for GPS interference mitigation.

**Conventional Beamformer:** In a conventional beamformer, the complex vector of weights  $W$  is equal to the array response vector  $a_c$ , which is defined by an array configuration (Moelker, 1996):

$$W_{conv} = a_c \quad (13)$$

For URA and UCA configurations the array response vector  $a_c$  is calculated using (5) and (8), respectively.

**MVDR Beamformer:** In GPS applications, the objective of adaptive beamforming is to maximize the gain in the direction of arrival of the desired signal from GPS and mitigate broadband interference incoming from the other directions. The weight vector  $W$  can be chosen to maximize the signal-to-interference-plus-noise ratio at the antenna output (Tummonery, 1994):

$$SINR = \frac{\sigma_s^2 |W^H a_c|^2}{W^H K_{j+n} W} \tag{14}$$

where  $K_{j+n}$  is the “interference + noise” covariance matrix of size  $(M \times M)$ , and  $\sigma_s^2$  is the signal power. The easy solution of the optimization problem (14) can be found by maintaining the distortionless response toward the desired signal and minimizing the power at the beamformer output. This criterion of optimization is formulated as:

$$\min_W W^H K_{j+n} W \quad \text{to subject} \quad W^H a_c = 1 \tag{15}$$

The solution of the optimization problem (15) is known as a minimum variance distortionless response beamformer (MVDR):

$$W_{MVDR} = \frac{K_{j+n}^{-1} a_c}{a_c^H K_{j+n}^{-1} a_c} \tag{16}$$

In practical applications, the covariance “interference plus noise” matrix  $K_{j+n}$ , is unavailable and the sample covariance matrix is used instead of it. The sample covariance matrix is estimated as:

$$\hat{K} = X^H X \tag{17}$$

Many practical applications of MVDR-beamformers require online calculation of the weights according to (16), and it means that the covariance matrix (17) should be estimated and inverted online. However, this operation is very computationally expensive and it may be difficult to estimate the sample covariance matrix in real time if the number of samples  $MN$  is large. Furthermore, the numerical calculation of the weights  $W_{MVDR}$  using the expression (16) may be very unstable if the sample covariance matrix is ill-conditioned. A numerically stable and computationally efficient algorithm can be obtained by using QR decomposition of the incoming signal matrix. The signal matrix is decomposed as  $X=QR$ , where  $Q$  is the unitary matrix and  $R$  is the upper triangular matrix. In that case the sample covariance matrix is calculated as:

$$\hat{K} = X^H X = (QR)^H (QR) = R^H R \tag{18}$$

Taking into account (18), the expression (16) takes the form:

$$W_{MVDR} = \frac{R(R^H)^{-1} a_c}{a_c^H R(R^H)^{-1} a_c} \tag{19}$$

In accordance with (19) the  $QR$ -based algorithm for calculation of beamformer weights includes the following three steps:

- The linear equation system  $R^H z_1 = a_c$  is solved. The solution is  $z_1^* = (R^H)^{-1} a_c$
- The linear equation system  $R z_2 = z_1^*$  is solved. The solution is  $z_2^* = R^{-1} z_1^*$
- The weight vector  $\hat{W}$  is obtained as  $\hat{W} = z_2^* / (a_c^H z_2^*)$

## 5.2 Circular cross-correlation

The main purpose of the cross-correlation stage is to identify the visible satellites in the incoming data and then find the beginning point of the C/A code and estimate the rough Doppler shift by correlating the incoming signal with the local signal replica. In a hardware-based GPS receiver, the conventional cross-correlation algorithm is implemented in hardware and the processing is performed in the time domain. In a software-based GPS receiver, the circular cross-correlation algorithm shown in Fig. 6 can be considered as a reduced computational version of the conventional cross-correlation method. In contrary to the conventional method the circular cross-correlation is performed in the frequency domain.

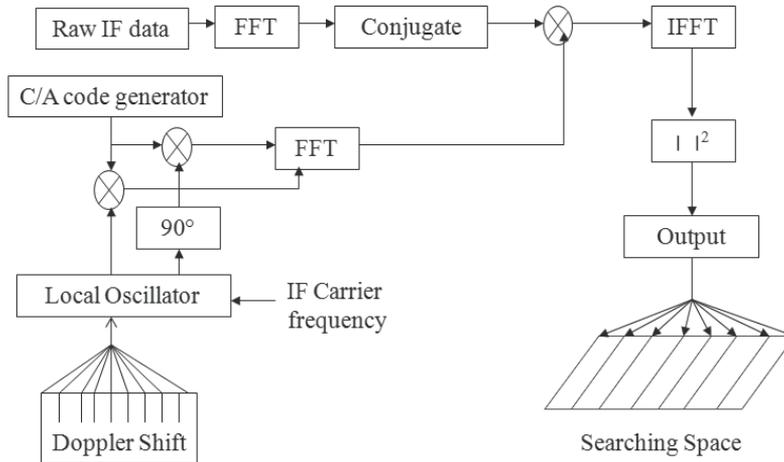


Fig. 6. Block diagram of the circular cross-correlation technique

At the circular cross-correlator output, the detection method should search over a frequency range of  $[-10\text{KHz}, 10\text{KHz}]$  to cover the expected Doppler frequency range for a high speed aircraft. This method is very suitable for software receiver implementation. The idea is that the input data sampled and stored in memory can be performed as blocks of data. Since the C/A code period is 1ms long, the circular cross-correlation must be performed on at least 1ms of the incoming data. According to (Tsui, 2005) the circular cross-correlation algorithm consists of the following steps:

- Perform the FFT of the input data  $x(k)$  converting them into frequency domain as  $X(k)$ .
- Take the complex conjugate  $X(k)$  obtaining the outputs  $X^*(k)$ .
- Generate 21 local codes  $l_{si}(k)$  as  $l_i(k) = c(k) \exp(2\pi j f_i)$ , where  $f_i = f_c + i \cdot \text{kHz}$ ,  $f_c$  is the intermediate frequency and  $i = -10, -9, \dots, 9, 10$ . The local code is the product of the C/A code of a satellite and a complex IF signal. The frequencies  $f_i$  of the local codes are separated by 1 kHz.

- Perform the FFT of the local codes  $l_{si}(k)$  to transform them to the frequency domain as  $L_i(k)$ .
- Multiply  $X^*(k)$  and  $L_i(k)$  point by point obtaining the result  $R_i(k)$ .
- Take the IFFT of  $R_i(k)$  to transform the result into time domain as  $r_i(k)$ .

### 5.3 CA CFAR detection

The constant false alarm rate (CFAR) signal detector is based on the principle of evaluating the noise power level to maintain the constant rate of false alarms (Finn & Johnson, 1968). The noise power estimate is calculated by integration of samples over a certain number of cells referred to as a reference window. The noise power estimate obtained is weighted with a scale factor to form an adaptive threshold, which is then compared to the signal from a cell under test. The structure of an adaptive Maximum Likelihood CFAR (ML CFAR) detector is shown in Fig. 7.

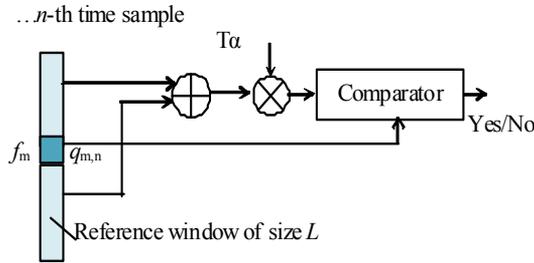


Fig. 7. Block diagram of the CFAR detector

In the ML CFAR detector, the decision rule for detection of the C/A code is formulated as:

$$\Phi_{m,n} = \begin{cases} 1, & \text{if } q_{m,n} \geq T_a \cdot w_{m,n} \\ 0, & \text{otherwise} \end{cases} \quad (20)$$

According to (20), if the test statistics  $q_{m,n}$  exceeds a threshold of detection  $T_a \cdot w_{m,n}$ , then the  $n^{\text{th}}$  time discrete and the  $m^{\text{th}}$  frequency bin give the beginning point ( $\tau_{\text{sat}}$ ) of the C/A code in  $(1/f_s)$  resolution in the input data and the carrier frequency ( $f_{c,\text{sat}}$ ) in 1kHz resolution:

$$f_{c,\text{sat}} = f_c + m \cdot 1\text{KHz}, \quad \tau_{\text{sat}} = n / f_s \quad (21)$$

The parameter  $f_s$  in (21) is the sampling frequency of the incoming IF data. The test statistics  $q_{m,n}$  in (20) is formed as:

$$q_{m,n} = \max_{i,k} \{r_i^2(k)\} \text{ for } i = m, k = n \quad (22)$$

The index  $i$  in (22) varies in the range of -10 to 10 and the index  $k$  varies from 1 to  $K$ , where  $K$  is the number of time samples of  $r_i$  on 1ms of data. The variable  $w_{m,n}$  in (20) is the estimate of the total noise power integrated over a reference window of length  $L$ . It is calculated as:

$$w_{m,n} = \sum_{\substack{l=m-L/2 \\ l \neq m}}^{m+L/2} r_m^2(n) \quad (23)$$

Since the sort procedure used for finding  $q_{m,n}$  in (22) is computationally expensive, we propose firstly to perform the thresholding over all the data in the search field with a single threshold and, secondly, to search the maximum only over the data that exceeded the threshold of detection. The decision rule of thresholding is formed as follows:

$$\Phi_{i,k} = \begin{cases} 1, & \text{if } r_i^2(k) \geq T_\alpha \cdot w \\ 0, & \text{otherwise} \end{cases} \quad (24)$$

In order to form a single threshold of detection ( $T_\alpha w$ ) in (24), we propose to use a reference window that includes  $L$  samples of  $r_i$  contained in four frequency bins, for example,  $f_{-10}=f_c-10\text{KHz}$ ;  $f_{-9}=f_c-9\text{KHz}$ ;  $f_{10}=f_c+10\text{KHz}$ ;  $f_9=f_c+9\text{KHz}$  (Fig. 8).

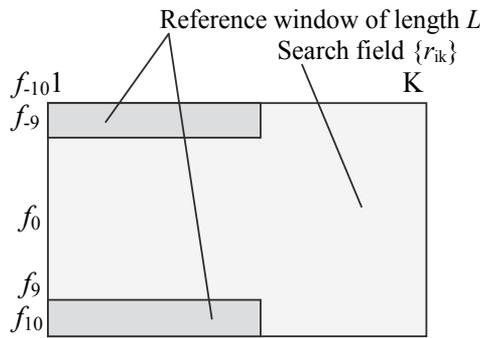


Fig. 8. Position of a reference window in the search field of data

In that case the noise power is estimated as:

$$w = \sum_{l=1}^{L/4} \{r_{-10}^2(l) + r_{-9}^2(l) + r_9^2(l) + r_{10}^2(l)\} \quad (25)$$

The time discrete  $n$  and the frequency bin  $m$ , which give the two C/A code parameters in (21), are determined as:

$$\{m, n\} = \arg \max \{r_i^2(k), \Phi_{i,k} = 1\} \quad (26)$$

The accuracy of the estimate  $w$  in (25) directly depends on the reference window length ( $L$ ). The choice of  $L$  depends on whether the reference window contains the GPS signal or not. There are three possible ways of choosing the appropriate window length:

- to use a small reference window of length  $L$  free of a GPS signal, which is enough to obtain the accurate estimate  $w$ ;
- to use a large reference window of length  $L_1$  ( $L_1 \gg L$ ) that contains a GPS signal where the influence of the GPS signal on the accuracy of  $w$  is insignificant;
- to use a small reference window of length  $L \ll L_1$  that contains a GPS signal where the maximal value is removed before estimation of the noise power.

The scale factor  $T_\alpha$  in (20 and 24) maintaining the required constant level of false alarm is:

$$P_{FA} = (1 + T_\alpha)^{-L} \quad (27)$$

## 6. Performance measures

The effectiveness of the GPS MDVR CFAR algorithm for detection of the C/A code in heavy noise environment is expressed in terms of three quality measures (Fig.4). These quality measures evaluate the influence of different factors on the performance of each processing stage.

### 6.1 SINR improvement factor

The performance of a beamforming algorithm can be evaluated in terms of the signal-to-interference-plus-noise ratio (SINR) improvement factor estimated at the beamformer output. For a single element array the input SINR is defined as:

$$SINR_{INPUT} = \frac{P_s}{P_N + P_i} \quad (28)$$

where  $P_s$  is the average power of the desired GPS signal,  $P_N$  is the receiver noise power, and  $P_i$  is the total broadband interference power. According to the superposition principle, the SINR at the beamformer output can be evaluated as

$$SINR_{OUT} = \frac{(W^H s)(W^H s)^H}{(W^H x_0)(W^H x_0)^H} \quad (29)$$

In (29),  $x_0$  is the total noise (noise + interference) at the antenna array element. The SINR improvement factor provided by the beamformer can be found as:

$$K_{SINR} = SINR_{OUT} / SINR_{INPUT} \quad (30)$$

As usual, this quality measure is expressed in dBs:

$$K_{SINR,dB} = SINR_{OUT,dB} - SINR_{INPUT,dB} \quad (31)$$

The SINR improvement factor evaluates the capability of the beamformer (conventional or MVDR) to cancel the interference power in the incoming signal and improve SINR at the beamformer output. If the Monte Carlo approach is used to estimate the SINR improvement factor, then the estimate of this quality measure is calculated as:

$$K_{SINR,dB} = \frac{1}{N_{total}} \sum_{n=1}^{N_{total}} K_{SINR,dB,n} \quad (32)$$

where  $K_{SINR,dB,n}$  is the estimate obtained in the n-th run, and  $N_{total}$  is the total number of Monte Carlo runs.

### 6.2 Post correlation SNR

The detectability of the C/A code directly depends on the signal-to-noise ratio at the cross-correlator output. The impact of the circular cross-correlation algorithm is evaluated in terms of the post correlation SNR estimated at the cross-correlator output:

$$SNR_{cor} = P_{max} / SLB_{ave} \quad (33)$$

where  $P_{\max}$  is the peak power at the cross-correlator output, and  $SLB_{\text{ave}}$  is the average sidelobe level. As usual, this quality measure is expressed in dBs:

$$SNR_{\text{cor,dB}} = P_{\max,\text{dB}} - SLB_{\text{ave,dB}} \quad (34)$$

If the Monte Carlo approach is used to estimate the post correlation SNR, then the estimate of this quality measure is calculated as:

$$SNR_{\text{cor,dB}} = \frac{1}{N_{\text{total}}} \sum_{n=1}^{N_{\text{total}}} SNR_{\text{cor,dB},n} \quad (35)$$

where  $SNR_{\text{cor,dB},n}$  is the estimate obtained in the  $n$ -th cycle of simulation

### 6.3 Probability of detection

This quality measure evaluates the capability of the overall three-stage GPS MDVR CFAR algorithm to detect the beginning point of the C/A code and find correctly the carrier frequency of the incoming IF signal while maintaining the required probability of false alarm. This performance measure can be evaluated using the Monte Carlo approach:

$$P_D = K_{\text{success}} / N_{\text{total}} \quad (36)$$

where  $K_{\text{success}}$  is the number of successful events, and  $N_{\text{total}}$  is the total number of runs.

## 7. Statistical analysis

### 7.1 Simulation algorithm

The performance of the GPS MDVR CFAR detection algorithm is evaluated using the Monte Carlo approach. The structure of the simulation algorithm is presented in Fig. 9.

The statistical estimates of the quality factors are evaluated for each stage of the detection algorithm. The goal of the Monte Carlo analysis is:

- to analyze the capability of two beamformers, non-adaptive (conventional) and adaptive (MVDR), to mitigate broadband radio frequency interference (RFI) at the navigation receiver input and as a result improve the post correlation SNR and increase the detectability of the C/A code in conditions of strong jamming;
- to evaluate the influence of several important factors on the performance of the joint three-stage detection algorithm. These factors are: interference-to-signal ratio ( $ISR$ ); planar array configuration ( $URA$  and  $UCA$ ); number of array elements ( $M$ ); sampling rate of the incoming data ( $f_s$ ); steering vector mismatch ( $\Delta\varphi, \Delta\theta$ ).

The simulation algorithm includes four consequential processing steps during each simulation cycle. Firstly, for each antenna element the complex valued signal is simulated in accordance with all important parameters of a GPS signal, interference and antenna array. The next processing steps include simulation of the performance of beamforming, cross-correlation and CFAR detection. For each simulation cycle, the current values of two quality measures, SINR improvement and SNR, are evaluated at the cross-correlator output by (31) and (34). If the beginning point and the carrier frequency of the C/A code are estimated correctly, then the counter  $K_{\text{success}}$  in (36) is incremented by 1. When all cycles of simulation are successfully accomplished, the statistical estimates of the three quality measures are calculated by (32), (35) and (36).

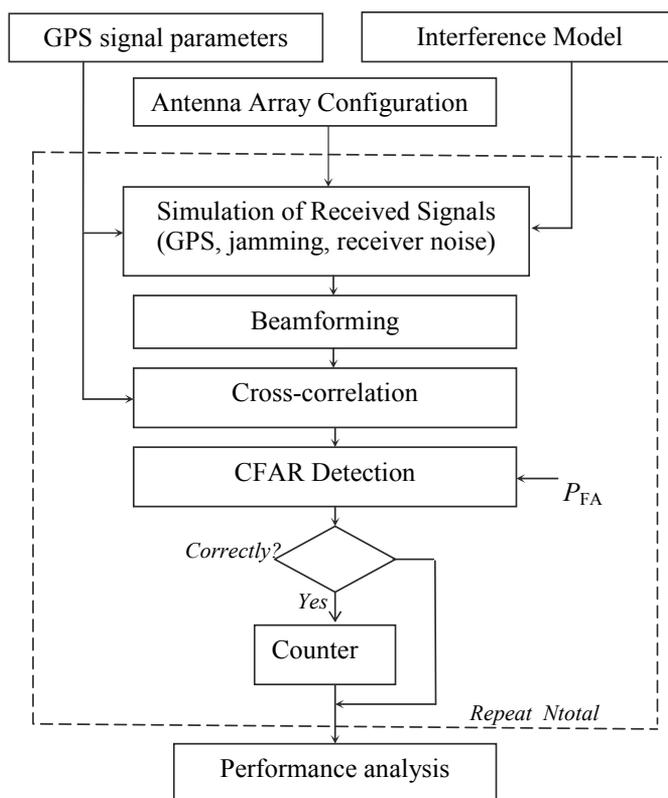


Fig. 9. Block diagram of the simulation algorithm

### 7.2 Simulation scenario

In this study, 1000 computer simulations of the adaptive three-stage algorithm for detection of the C/A code are performed in order to evaluate the influence of such factors as array configuration, number of array elements and sampling rate on the capability of this detection algorithm to operate effectively in conditions of strong jamming. In GPS applications, the interelement spacing in antenna arrays is approximately 0.09m for frequency L1, and this technical demand puts a physical limitation on how small the array can be used in a GPS receiver and how many elements are appropriate to be used in such an antenna array. We consider in the study two small array configurations shown in Fig. 10.

In the first antenna array (UCA-7), the first (reference) element is located at the array center but the other six elements are circular relative to the center. The second antenna array (URA-9) is rectangular and contains nine elements with half-wavelength interelement spacing. Both antenna arrays have the same overall dimensions of about 20cm. It is well known that the number of antenna elements  $M$  is related to the number of broadband jammers that can be nulled by the beamforming algorithm. Typically, the number of broadband jammers that can be nulled by the beamforming algorithm corresponds to  $(M-1)$ . Taking into account the angular resolution of the two antenna arrays, the simulation scenario includes the presence of one satellite and four jamming sources located at different positions in space. The angular coordinates of both, satellite and jammers, are presented in Table 1.

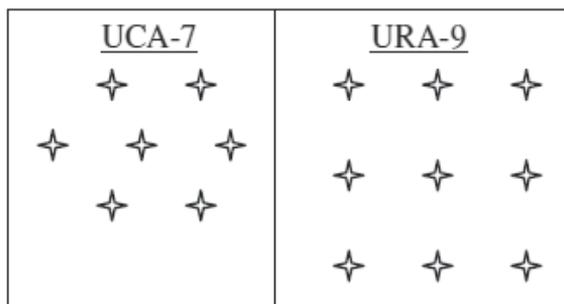


Fig. 10. Antenna array configurations used in simulations

Variant	Jamming	GPS signal
IF carrier: 1.2513 MHz Sampling: 5.0053 MHz	Four jammers: Elevation: $\theta=40^\circ$ Azimuth: $\varphi_1=-70^\circ$ ; $\varphi_2=-60^\circ$ ; $\varphi_3=60^\circ$ ; $\varphi_4=70^\circ$	Elevation: $\theta=40^\circ$ Azimuth: $\varphi=0^\circ$ Doppler shift: 5 kHz SNR: -20dB
IF carrier: 2.4967 MHz Sampling: 9.9868 MHz	ISR: 10dB ... 100dB	Duration: 1ms C/A code: satellite 19

Table 1. Inteference and signal parameters

According to (10), the intensity of the signal received from a satellite is determined by the value of the signal-to-noise ratio (SNR) observed at the receiver input. In real situations, the SNR observed at the receiver input is between -15 and -30 dB. In our simulations, the SNR is assumed to be equal to -20dB. In (11), the interference intensity is determined by the value of the interference-to-noise ratio (INR) observed at the receiver input. In our case the parameter INR is expressed in terms of the interference-to-signal ratio (ISR), which describes the severity of the interference situation at the GPS receiver input.

$$INR_{dB} = ISR_{dB} + SNR_{dB} \quad (37)$$

In simulations, the parameter ISR varies between 10dB and 100 dB in increments of 5dB.

### 7.3 Simulation results

In broadband interference environment, the effectiveness of the GPS MDVR CFAR detection algorithm depends on the capability of the beamforming stage as much as possible to suppress interference before the cross-correlation stage. In order to compare the effectiveness of the two beamforming algorithms, the SINR improvement factor is evaluated at the output of each of the two beamformers (conventional and MVDR) as a function of the input ISR. Figure 11(a) shows that the capability of each beamformer to suppress broadband interference depends on the array geometry (URA and UCA) and the number of array elements as well (9 and 7). It can be seen that the MVDR-beamformer very successfully mitigates broadband interference even if the interference intensity becomes 100 dB over the desired GPS signal. The MVDR-beamformer is the most effective for URA-9. The results presented in Fig.11(a) confirm that in contrast to the conventional beamformer the adaptive MVDR algorithm is able to almost completely to suppress broadband interference. Because the precise angular location of the GPS satellite is not always known, it is very important to

analyze the sensitivity of each beamformer to angular errors in satellite location (in both azimuth and elevation). The sensitivity of each beamformer to angular (azimuth and elevation) errors in satellite location is shown in Fig. 11(b) for  $ISR=100$ dB. Close inspection of the results presented in Fig.11(b) reveals that in the case of the MVDR beamformer the losses in SINR due to steering vector mismatch are tolerable when angular errors are in the interval of  $[-15^\circ, 15^\circ]$ . However, in the conventional beamformer the losses in SINR drastically rise with increase of angular errors in satellite location.

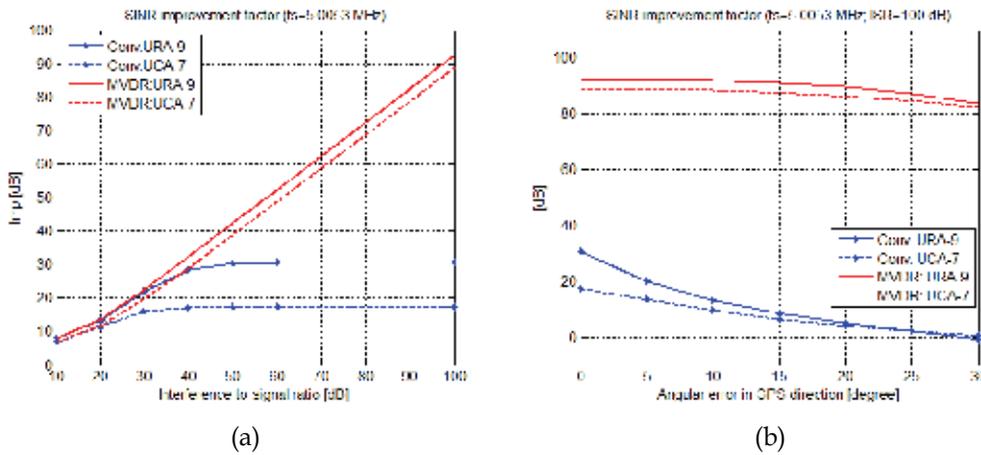


Fig. 11. SINR improvement factor as a function of ISR in case of: a)- zero angular errors , b)- non-zero angular errors in satellite position

For such a beamformer, when angular errors in satellite location equal to  $5^\circ$ , the SINR losses exponentially increase to the levels of 20dB – for URA-9, and 5dB – for UCA-7. The effectiveness of the first two processing blocks of the GPS MDVR CFAR detection algorithm can be expressed in terms of the SNR estimated at the cross-correlator output (Fig.12).

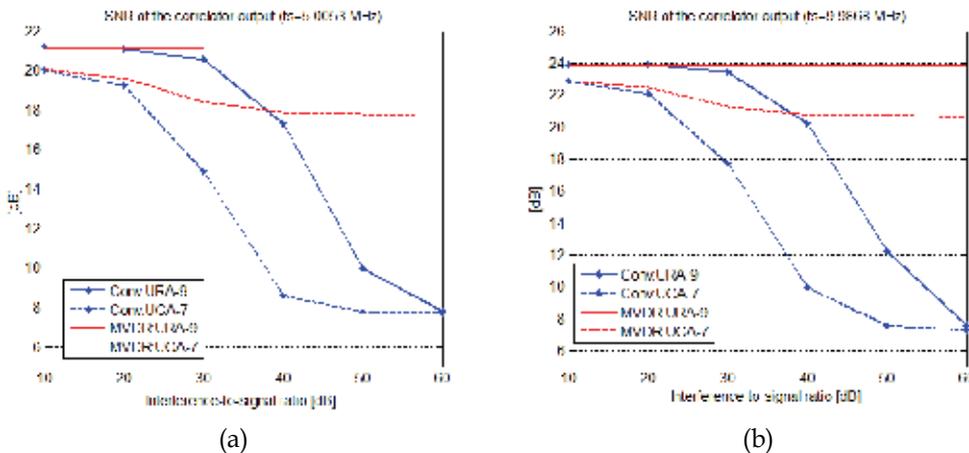


Fig. 12. SNR at the cross-correlator output as a function of ISR evaluated for two sampling frequencies : (a) -  $f_s=5.0053$  MHz), (b) -  $f_s=9.9868$  MHz)

For a stable GPS signal tracking, the post correlation SNR must be at least 20 dB (Tsui, 2005). Higher post correlation SNR means the higher probability of an event in which the GPS signal is detected in the incoming signal and its both parameters, the beginning point and the carrier frequency, are correctly estimated. The results illustrated in Fig. 12 clearly show that the post correlation SNR quickly degrades with increase of the broadband interference intensity, i.e. ISR. Comparison analysis of the results from Fig. 12(a) and Fig. 12(b) shows that the post correlation SNR rises with increase of the sampling rate of the incoming data. When the sampling frequency is 10 MHz both antenna array configurations, URA-9 and UCA-7, can guarantee post correlation SNR above 20dB even if the interference intensity becomes 60 dB over the desired GPS signal. However, when the sampling frequency is 5 MHz, only URA-9 provides the SNR above 20 dB in the whole range of ISR.

The results shown in Fig. 13 demonstrate the sensitivity of the cross-correlation performance to steering vector mismatch. The results are obtained for two values of the sampling rate: Fig. 13(a) - for 5 MHz and Fig. 13(b) - for 10 MHz. The study is performed for  $ISR = 100$  dB. It can be seen that when sampling frequency is 5 MHz, only the rectangular antenna array with 9 elements can guarantee the post correlation SNR above 20dB provided, however, that angular errors in satellite location do not exceed  $15^\circ$ . When the sampling frequency is doubled, both arrays can guarantee the post correlation SNR of above 20dB if angular errors in satellite location are less than  $10^\circ$  - for UCA-7 and less than  $22^\circ$  - for URA-9. These graphical results also suggest that the influence of angular errors in satellite location can be partially compensated by increasing the sampling rate of the incoming data.

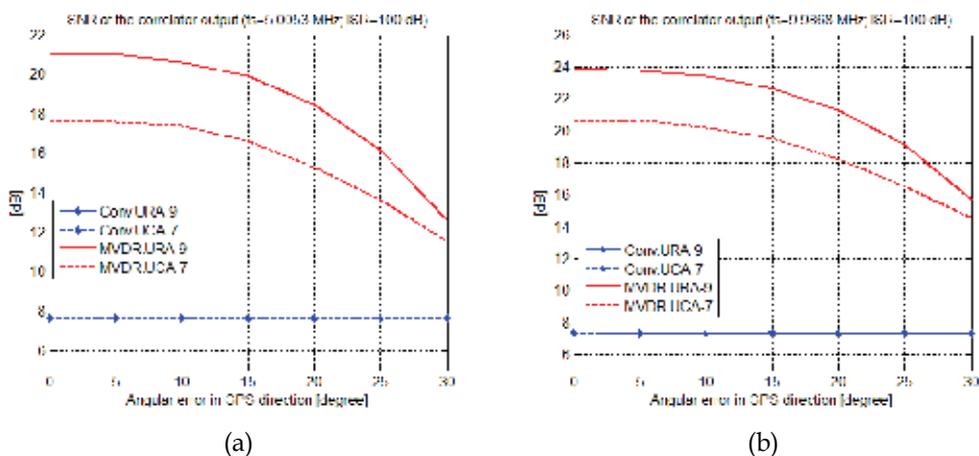


Fig. 13. SNR at the cross-correlator output as a function of angular errors evaluated for two sampling frequencies : (a) -  $f_s = 5.0053$  MHz), (b) -  $f_s = 9.9868$  MHz)

The CFAR thresholding procedure is simulated using the modified test statistics (26). The noise level estimate  $w$  is formed using the expression (25). The scale factor  $T_\alpha$  in (27) is determined to maintain the required probability of false alarm ( $P_{fa} = 10^{-7}$  and  $10^{-9}$ ). Since in real situations the presence of a GPS signal in the reference window is unknown, the objectives of the study are: (i)-to determine the minimal window of length  $L_1$ , in which the influence of the presence of a GPS signal on the accuracy of the estimate  $w$  is insignificant; (ii)- to compare two CFAR algorithms, the first of which uses the reference window of

length  $L_1$  and the other uses the reference window of length  $L$  ( $L_2 \ll L_1$ ) where the maximal value is removed from the reference window before estimation of the noise level. In Fig.14, the probability of detection is plotted as a function of the interference-to-signal ratio when two beamforming algorithms are used to mitigate broadband interference at the receiver input. These results are presented for different reference windows free of a GPS signal: Fig.14(a) - for the conventional beamforming algorithm and Fig.14(b) - for the MVDR. The detection probabilities are evaluated for  $P_{fa} = 10^{-9}$ . These results confirm that in case when a GPS signal is not present in the reference window, the minimal length of a reference window needed to accurately estimate the noise level is  $L=60$ . It can be seen that unlike the conventional beamformer the MVDR-beamformer successfully mitigates broadband interference and therefore the GPS MVDR CFAR detection algorithm maintains the high probability of detection within a wide diapason of the intensity of interference (ISR).

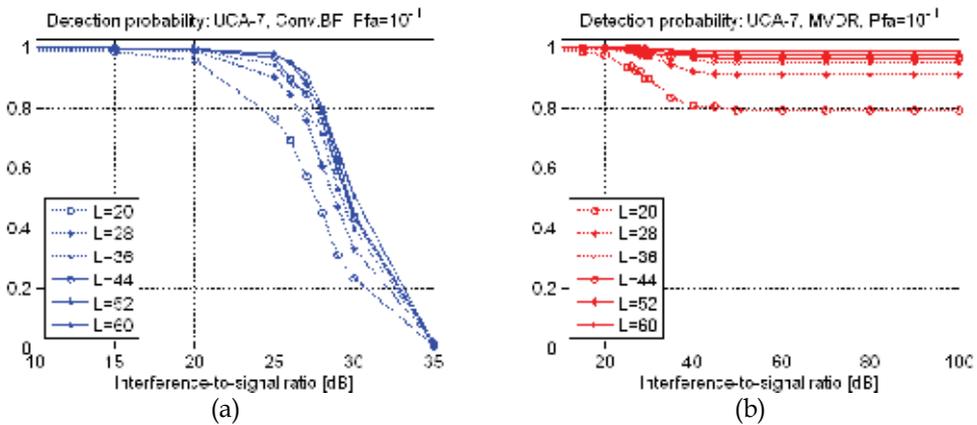


Fig. 14. Detection probability evaluated for reference windows free of a satellite signal: (a) - conventional BF ; (b) - MVDR

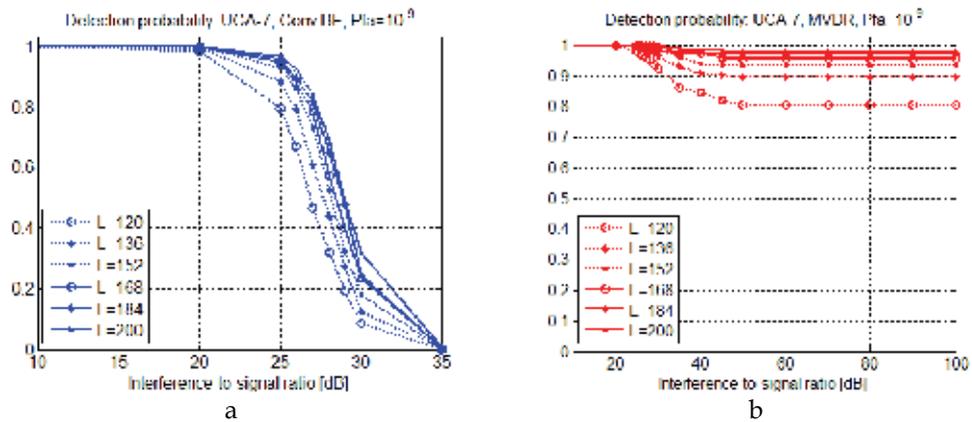


Fig. 15. Detection probability evaluated for reference windows with the GPS signal: (a) - conventional BF; (b) - MVDR

In Fig. 15, the probability of detection is plotted as a function of ISR when the reference window used for estimation of the noise level contains a GPS signal. The reference window

length varies between 120 and 200 elements. The detection probabilities evaluated for  $P_{fa}=10^{-9}$  are shown in Fig. 15(a) - for the conventional beamforming algorithm and Fig. 15(b) - for the MVDR algorithm. It is shown that a 200-element reference window is sufficient for accurate estimation of the noise level when it contains a GPS signal ( $L_1=200$ ).

The detection probabilities plotted in Fig.16 are obtained in case when a 61-element reference window that contains a GPS signal is used for estimation of the noise level. The results illustrate that this reference window can be used for estimation of the noise power if its maximal sample is previously removed before estimation (Conv.BF-OS and MVDR-OS).

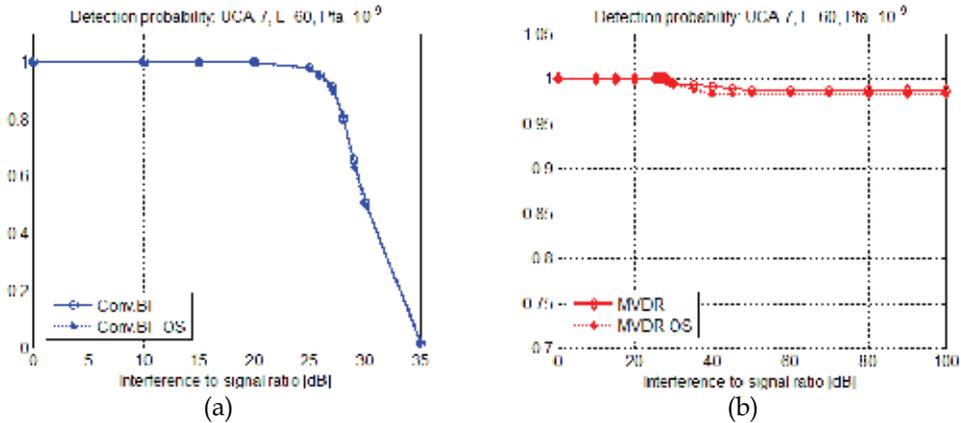


Fig. 16. Detection probability as a function of ISR evaluated for a 60-element reference window free of a GPS signal (Conv.BF and MVDR) and for a 61-element reference window containing a GPS signal (Conv.BF-OS and MVDR-OS)

The results shown in Fig. 17 demonstrate the sensitivity of the detection performance to steering vector mismatch. In Fig. 17(a), the detection probability is plotted for two beamforming algorithms, two values of  $P_{fa}$  and  $ISR=25$ dB. The analysis of the results shows that the probability of detection degrades with increase of angular errors in satellite location.

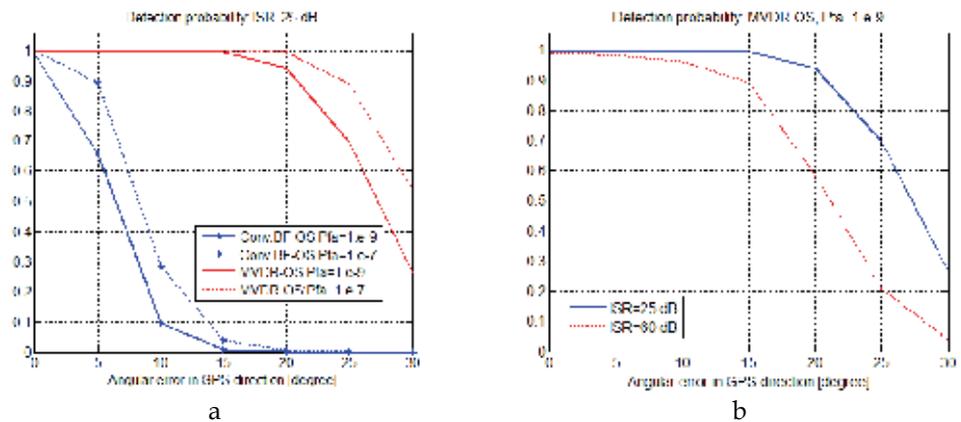


Fig. 17. Detection probability as a function of angular errors evaluated for a 61-element reference window with removal of its maximal value before estimation of the noise level: a)  $ISR=25$  dB,  $P_{fa}=10^{-7}$  and  $10^{-9}$ ; b)  $ISR=25$  dB and 60 dB

The results presented in Fig. 17(b) and obtained for two values of ISR show that the detection performance becomes more sensitive to steering vector mismatch when the interference intensity increases.

Finally, it can be concluded that the GPS MDVR CFAR detection algorithm can guarantee the detection probability of above 0.9 while maintaining the false alarm probability of  $10^{-9}$  when angular errors in satellite location do not exceed  $15^\circ$  even if the interference-to-signal ratio becomes very high (60 dB).

## 8. Conclusion

The proposed three-stage GPS MDVR CFAR algorithm for detection of the C/A code in the incoming IF data has very good antijamming protection in a wide range of interference intensity. The obtained results reveal show that the GPS MDVR CFAR detection algorithm can guarantee the high detection probability while maintaining the required probability of false alarm even if the interference-to-signal ratio becomes very high.

## 9. Acknowledgment

This work is supported by the Bulgarian Science Fund (the project DTK02/28-17.12.2009) and SISTER, FP7-REGPOT-2007-1

## 10. References

- Behar, V.; Kabakchiev, Ch. & Rohling, H. (2010). MVDR Radar Signal processing Approach for Jamming Suppression in Satellite Navigation Receivers, *Proc. of the 11-th Intern. Radar Symp. IRS-2010*, pp.134-137, ISBN 978-9955-690-18-4, Vilnius, Lithuania, June 2010.
- Behar, V.; Kabakchiev, Ch. & Rohling, H. (2010). MVDR Beamformer with a CFAR Processor for Jamming Suppression in GPS Receivers, *Proc. of the Intern. Symp. on Radio Systems and Space Plasma ISRSSP'10*, pp.9-12, ISBN 978-989-8425-27-0, Sofia, Bulgaria Aug. 2010.
- Behar, V.; Kabakchiev, Ch.; Gaydadjiev, G.; Kuzmanov, G. & Ganchosov, P. (2009). Parameter Optimization of the Adaptive MVDR QR-based Beamformer for Jamming and Multipath Suppression in GPS/GLONASS Receivers, *Proc. of the 16<sup>th</sup> International Conference on Integrated Navigation Systems, ICINS'2009*, pp. 325-334, ISBN 978-5-900780-69-6, Saint-Petersburg, Russia, May, 2009.
- Behar, V.; Vassileva, B. & Kabakchiev, Ch. (2010). A Simulation Tool for Analysis of MTD Algorithms Employing STAP Techniques, *Proc. of the 11-th Intern. Radar Symp. IRS-2010*, pp. 516-519, ISBN 978-9955-690-18-4, Vilnius, Lithuania, June 2010.
- Behar, V.; & Kabakchiev Ch. (2009). Multiple Signal Extraction in Jamming using Adaptive Beamforming with Arbitrary Array Configurations, *Cybernetics and Information Technologies*, Vol. 9, No. 3, 2009, pp. 76-85 , ISSN: 1311-9702.
- Behar V.; Kabakchiev Ch. & Doukovska, L. (2000). Adaptive CFAR PI Processor for Radar Target Detection in Pulse Jamming, *Journal of VLSI Signal Processing*, Vol. 26, 2000, pp. 383 - 396, ISSN:0922-5773.
- Finn, H. & Johnson, R. (1968). Adaptive detection mode with threshold control as a function of spatially sampled clutter estimation, *RCA Review*, Vol. 29, No. 3, 1968, pp. 414-464, ISSN: 0033-6831.

- Fu, Z.; Hombostel, A.; Hammesfahr, J. & Konovaltsev A. (2003). Suppression of multipath and jamming signals by GPS/Galileo applications, *GPS Solutions*, No. 6, 2003, pp. 257-264, ISSN: 1080-5370.
- Ganchosov, P.; Kuzmanov, G.; Kabakchiev H.; Behar V.; Romansky, R. & Gaydadjiev G. (2009). FPGA Implementation of Modified Gram-Schmidt QR-Decomposition, *Proceedings of the 3rd HiPEAC Workshop on Reconfigurable Computing*, pp. 41-51, Paphos, Cyprus, January 2009.
- Garvanov, I.; Behar V. & Kabakchiev Ch. (2003). CFAR Processors in Pulse Jamming, *LNCS*, Vol. 2542, 2003, pp. 291-298, ISSN: 0302-9743.
- Ioannides, P. & Balanis, C. (2005) "Uniform circular and rectangular arrays for adaptive beamforming applications", *IEEE Trans. Antenn. Wireless Propagat. Lett.*, vol.4., 2005, pp. 351-354, ISSN: 1536-1225.
- Kabakchiev, Ch.; Behar, V. & Rohling, K. (2010). Adaptive C/A Code Acquisition in Conditions of Broadband Interference with MVDR and CFAR Techniques, *Proc. of the European Navig. Conf. on Glob. Navig. Satel. Syst. - ENC GNSS 2010*, Braunschweig, Germany, Oct. 2010 (accepted).
- Kabakchiev, Ch.; Rohling, H.; Garvanov, I.; Behar V. & Kyovtorov, V. (2010). Multisensor Detection in Randomly Arriving Impulse Interference using the Hough Transform, In: *Radar Technology*, Guy Kouemou (Ed.), pp. 179-204, In-Teh, ISBN: 978-3-902613-49-3), Vukovar, Croatia, 2010.
- Kabakchiev, Ch.; Behar, V.; Rohling, H.; Garvanov, I.; Kyovtorov, V. & Kabakchieva, D. (2010) Analysis of Multi-Sensor Radar Detection based on the TBD-HT Approach in ECM Environment, *Proc. of the IEEE Radar Conference - RADAR'10*, pp. 651-656, ISBN 978-1-4244-5813-4, Washington DC, USA, May 2010.
- Moelker, D. (1996). Adaptive antenna arrays for interference cancellation in GPS and GLONASS receivers, *Proc. of the IEEE conf. on Position Location and Navigation*, pp. 191-196, ISBN: 0-7803-3085-4, Atlanta, GA, Apr. 1996.
- Sklar, J. (2003). Interference mitigation approaches for the Global Positioning System, *MIT Lincoln Laboratory Journal*, vol.14, No 2, 2003, pp. 167-177
- Soubielle,J.; Fijalkow, I.; Duvau, P. & Bibaut, A. (2002). GPS positioning in a multipath environment, *IEEE Trans. on Signal Processing*, Vol. 50, No 1, 2002, pp. 141-150, ISSN: 1053587X
- Tsui,J. (2005). *Fundamentals of Global Positioning System Receivers: A Software Approach*, John Wiley & Sons Inc, ISBN: 0-471-70647-7, New York
- Tummonery,L.; Proudler, I.; Farina, A. & McWhirter, J. (1994). QRD-based MVDR algorithm for adaptive multi-pulse antenna array signal processing, *IEE Proc. Radar, Sonar, Navigation*, Vol. 141, No. 2, 1994, pp. 93-102, ISSN: 1350-2395.
- van Trees, H. (2002). *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*, JohnWiley and Sons Inc., ISBN: 978-0-471-09390-9, New York
- Vouras, P. & Freburger, B. (2008). Application of adaptive beamforming techniques to HF radar, *Proc. of the IEEE conf. RADAR'08*, pp.6, ISSN: 1097-5659, Rome, Italy, May 2008

# Practical Monte Carlo Based Reliability Analysis and Design Methods for Geotechnical Problems

Jianye Ching  
*National Taiwan University, Taiwan,  
Republic of China*

## 1. Introduction

Reliability analysis is an important tool for quantifying uncertainties in analysis and design of engineering systems. In the past decades, the so-called first-order reliability method (FORM) (Ang & Tang, 1984) was the main stream method for reliability analysis. This method transforms a reliability analysis problem into an approximate optimization problem so that the required computation is minimized. Nonetheless, such transformation comes with some premises and tradeoffs: (a) to make the optimization problem tractable, the number of random variables of the target problem cannot be too many; (b) the problem at hand is better to be lightly nonlinear to avoid large bias in the estimated reliability; and (c) the engineers must have basic skills for solving nonlinear optimization problems.

The first two premises may be questionable for realistic geotechnical problems because there are typically numerous random variables in realistic geotechnical engineering analyses and designs. Although techniques are developed to reduce the number of random variables (e.g., Ghanem & Spanos, 1991), their generality and accuracy are not yet proved. Therefore, for realistic geotechnical engineering analyses and designs, FORM may not be the best solution. More seriously, average engineers may not have the knowledge and skills for nonlinear optimization. It is not trivial for them to implement FORM, even for the simplest geotechnical design examples.

Given the rapid growth of nowadays personal computers (PCs), massive computations are now more possible than ever. In particular, Monte Carlo simulations (MCS) can nowadays be implemented for the purpose of reliability analyses even with PCs. MCS is general for the number of random variables and the problem complexity; hence the limitation of FORM can be easily overcome. Moreover, the basic idea of MCS is very simple and intuitive. Finally, geotechnical models can be treated as black boxes when implementing MCS. All these features make MCS attractive for practicality. The only criticism for MCS is that it is inefficient for problems with very small failure probabilities (or with very high reliabilities). However, this limitation has been gradually removed by the recent advancements in the Monte Carlo based reliability methods.

The goal of this chapter is to demonstrate the uses of some Monte Carlo based reliability methods and reliability-based design methods. In particular, a realistic geotechnical design example is developed for the purpose of demonstration: the implementation of all methods

will be presented based on the same example. First of all, this chapter will review practical Monte Carlo based reliability analysis methods, including

- a. Direct Monte Carlo simulation
- b. Importance sampling
- c. Subset simulation

The traditional FORM will be also briefly reviewed for completeness.

Second, this chapter will review state-of-the-art developments in the Monte Carlo based reliability-based design methods. This subject is the inverse problem of reliability analyses: the purpose of reliability analyses is to obtain the reliability given the design dimension of an engineering system, but the purpose of reliability-based design is to obtain the design dimension given the target reliability. The author himself (Ching & Phoon, 2010) has developed a series of Monte Carlo based methods in this line. The review will be limited to the following design methods:

- a. Monte Carlo based safety factor design
- b. Monte Carlo based load-resistance factor design
- c. Monte Carlo based multiple resistance factor design
- d. Monte Carlo based partial factor design

As opposed to the FORM-based reliability-based design methods, these Monte Carlo based methods are, again, not limited by the number of random variables and problem complexity and not requiring the acknowledge of optimization skills.

## 2. Design example for demonstration

Throughout this chapter, an example of geotechnical designs will be given to demonstrate the reviewed reliability analysis and reliability-based design methodologies. Consider a drilled shaft of 74.7 m long that is to be built at a site with ground profile shown in Table 1,

Type	Depth (m)	Middle depth $d$ (m)	$\sigma'_{v,m}$ (kN/m <sup>2</sup> )	Thickness $t$ (m)	$s_{u,m}$ (kN/m <sup>2</sup> )	$q_{u,m}$ (kN/m <sup>2</sup> )
Clay	0.0 - 42.1			$42.1 = t_c$	70	
Sand	42.1 - 49.6	$45.85 = d_s$	$350 = \sigma'_{vs,m}$	$7.5 = t_s$		
Gravel	49.6 - 69.7	$59.65 = d_g$	$480 = \sigma'_{vg,m}$	$20.1 = t_g$		
Sandstone	69.7 - 74.7			$5.0 = t_r$		900

Table 1. Ground profile for the example design site

where there are four strata, including clay 'c', sand 's', gravel 'g', and rock 'r' layers;  $d_x$  and  $t_x$  respectively denote the middle depth and thickness of each layer (the subscript 'x' may be either 'c', 's', 'g', or 'r', depending on the associated stratum type);  $\sigma'_{vs,m}$ ,  $\sigma'_{vg,m}$ ,  $s_{u,m}$  and  $q_{u,m}$  are respectively the measured in-situ effective stress in sand layer, in-situ effective stress in gravel layer, undrained shear strength of clay layer, and uniaxial compression strength of rock layer. The measurement is subjected to measurement errors:

$$\begin{aligned} \ln(\sigma'_{vs,m}) &= \ln(\sigma'_{vs}) + e_{\sigma'_{vs}} & \ln(\sigma'_{vg,m}) &= \ln(\sigma'_{vg}) + e_{\sigma'_{vg}} \\ \ln(s_{u,m}) &= \ln(s_u) + e_{s_u} & \ln(q_{u,m}) &= \ln(q_u) + e_{q_u} \end{aligned} \quad (1)$$

where  $\sigma'_{vs}$ ,  $\sigma'_{vg}$ ,  $s_u$ ,  $q_u$  are the corresponding actual values, and  $e_{\sigma'_{vs}}$ ,  $e_{\sigma'_{vg}}$ ,  $e_{s_u}$ ,  $e_{q_u}$  quantify measurement errors. These measurement errors are modeled as zero-mean normal random variables with standard deviations chosen to be 0.1, 0.1, 0.2, and 0.5, respectively.

The axial compression capacity (Q) of the drilled shaft is provided by the side resistance (S) and tip resistance (T), and it can be computed using the equation given below:

$$Q = S + T \quad (2)$$

Although the shaft tip may contribute to the overall compression resistance, the majority of the compression capacity is provided by the side resistance. Its contribution to the overall capacity often is ignorable compared to the side resistance. The side resistance is provided by the shaft adhesion for cohesive soils and rocks and shaft frictional resistance for cohesionless soils:

$$S = S_c + S_s + S_g + S_r \quad (3)$$

where  $S_c$ ,  $S_s$ ,  $S_g$ , and  $S_r$  are side resistances for the clay, sand, gravel, and rock layers, respectively. The side resistance in a given layer, denoted by  $S_x$ , can be computed as

$$S_x = \pi B f_{sx} t_x \quad (4)$$

where B is the diameter of the shaft;  $f_{sx}$  is the unit side resistance provided by layer 'x'. The unit side resistance  $f_s$  is correlated to geotechnical parameters such as  $s_u$ ,  $\sigma'_{vs}$ , and  $q_u$ . Useful empirical correlation equations are listed in Table 2, where  $\epsilon$ 's quantify the transformation uncertainties. These  $\epsilon$ 's are modeled as zero-mean normal random variables with standard deviations listed in the table. It is then clear that

$$\begin{aligned} S_c &= \pi B e^{2.7+0.3 \ln s_u + \epsilon_{S_c}} t_c = \pi B e^{2.7+0.3 [\ln s_{u,m} - e_{s_u}] + \epsilon_{S_c}} t_c \\ S_s &= \pi B e^{1.0802 - 0.6588 \ln(d_s) + \ln(\sigma'_{vs}) + \epsilon_{S_s}} t_s = \pi B e^{1.0802 - 0.6588 \ln(d_s) + [\ln(\sigma'_{vs,m}) - e_{\sigma'_{vs}}] + \epsilon_{S_s}} t_s \\ S_g &= \pi B e^{2.1792 - 0.7528 \ln(d_g) + \ln(\sigma'_{vg}) + \epsilon_{S_g}} t_g = \pi B e^{2.1792 - 0.7528 \ln(d_g) + [\ln(\sigma'_{vg,m}) - e_{\sigma'_{vg}}] + \epsilon_{S_g}} t_g \\ S_r &= \pi B e^{3.0253 + 0.414 \ln q_u + \epsilon_{S_r}} t_r = \pi B e^{3.0253 + 0.414 [\ln q_{u,m} - e_{q_u}] + \epsilon_{S_r}} t_r \end{aligned} \quad (5)$$

Reliability analyses and reliability-based designs will be demonstrated on this design example. The drilled shaft is subjected to an axial dead load  $L_D$  and axial live load  $L_L$ . They are modeled as lognormal random variables with mean values  $\{\mu_{LD}, \mu_{LL}\}$  and coefficients of variation (c.o.v.)  $\{\delta_{LD} = 0.1, \delta_{LL} = 0.25\}$ . The herein goal is to demonstrate (a) the calculation of the reliability of a drilled shaft with given dimension (i.e., diameter  $B = 1.2$  m and length  $L = 74.7$  m) and to demonstrate (b) the determination of the required dimension B and L to achieve a prescribed target reliability. The item (a) is the goal for reliability analysis, while item (b) is for reliability-based design.

The collection of random variables is denoted by  $X \in \mathbb{R}^p$ , where p is the dimension of X. For this example, X includes the measurement errors  $\{e_{\sigma'_{vs}}, e_{\sigma'_{vg}}, e_{s_u}, e_{q_u}\}$ , transformation uncertainties  $\{\epsilon_{S_c}, \epsilon_{S_s}, \epsilon_{S_g}, \epsilon_{S_r}\}$  and loads  $\{L_D, L_L\}$ . The collection of design parameters is denoted by  $\theta \in \mathbb{R}^q$ , where q is the dimension of  $\theta$ . For this example,  $\theta$  includes the diameter B and shaft length L. Let F denotes the failure event:  $F = \{SR(X, \theta) < 1\}$ , where  $SR(X, \theta)$  is called the

safety ratio, the random version of the classical safety factor. In general, a safety ratio less than 1 does not necessarily imply the complete collapse of the system but does imply unsatisfactory performance of the system in the sense of violating some limit states, e.g. serviceability, repairable, or ultimate limit states. Throughout the chapter, it can be assumed without loss of generality that  $SR(X,\theta)$  is positive and that the probability density function (PDF) of the random vector  $X$  conditioned on  $\theta$  (i.e.,  $\theta$  takes specific numerical values), denoted by  $p(x|\theta)$ , is known.

	Correlation Model for Unit Side Resistance $f_s$ (kN/m <sup>2</sup> )	Standard Deviation of Transformation Uncertainty $\varepsilon$
Clay	$f_s = \exp(2.7 + 0.3\ln(s_u) + \varepsilon_{S_c})$	0.3216
Sand	$f_s = \exp(1.0802 - 0.6588\ln(d) + \ln(\sigma'_v) + \varepsilon_{S_s})$	0.5414
Gravel	$f_s = \exp(2.1792 - 0.7528\ln(d) + \ln(\sigma'_v) + \varepsilon_{S_g})$	0.6689
Rock	$f_s = \exp(3.0253 + 0.414\ln(q_u) + \varepsilon_{S_r})$	0.7160

Table 2. Correlation models for evaluating unit side resistance and the associated uncertainty for various strata

For this particular example, the safety ratio  $SR(X,\theta)$  can be defined as:

$$SR = \frac{S_c + S_s + S_g + S_r}{L_D + L_L} \tag{6}$$

As will be clear later, it is convenient to transform the entire problem into the standard Gaussian space, i.e.,

$$SR(Z,\theta) = \frac{\pi B \left[ e^{2.7+0.3[\ln s_{u,m}-0.2z_{su}]+0.3216z_{Sc}} t_c + e^{1.0802-0.6588\ln(d_s)+[\ln(\sigma'_{vs,m})-0.1z_{\sigma'_{vs}}]+0.5414z_{Ss}} t_s \right.}{+e^{2.1792-0.7528\ln(d_g)+[\ln(\sigma'_{vg,m})-0.1z_{\sigma'_{vg}}]+0.6689z_{Sg}} t_g + e^{3.0253+0.414[\ln q_{u,m}-0.5z_{qu}]+0.7160z_{Sr}} t_r \left. \right]}{e^{\ln(\mu_{LD}/\sqrt{1+0.1^2})+\sqrt{\ln(1+0.1^2)}\cdot z_{LD}} + e^{\ln(\mu_{LL}/\sqrt{1+0.25^2})+\sqrt{\ln(1+0.25^2)}\cdot z_{LL}}} \tag{7}$$

$$= \frac{\pi B \left[ e^{2.7+0.3[\ln(70)-0.2z_{su}]+0.3216z_{Sc}} \cdot 42.1 + e^{1.0802-0.6588\ln(45.85)+[\ln(350)-0.1z_{\sigma'_{vs}}]+0.5414z_{Ss}} \cdot 7.5 \right.}{+e^{2.1792-0.7528\ln(59.65)+[\ln(480)-0.1z_{\sigma'_{vg}}]+0.6689z_{Sg}} \cdot 20.1 + e^{3.0253+0.414[\ln(900)-0.5z_{qu}]+0.7160z_{Sr}} \cdot 5.0 \left. \right]}{e^{\ln(\mu_{LD}/\sqrt{1+0.1^2})+\sqrt{\ln(1+0.1^2)}\cdot z_{LD}} + e^{\ln(\mu_{LL}/\sqrt{1+0.25^2})+\sqrt{\ln(1+0.25^2)}\cdot z_{LL}}}$$

where  $Z = \{z_{\sigma'_{vs}}, z_{\sigma'_{vg}}, z_{su}, z_{qu}, z_{Sc}, z_{Ss}, z_{Sg}, z_{Sr}, z_{LD}, z_{LL}\}$  are jointly standard Gaussian random variables, i.e.,

$$p(z|\theta) = \frac{1}{2\pi^{p/2}} e^{-\frac{1}{2}z^T z} \tag{8}$$

$\{\sigma'_{vs,m}, \sigma'_{vg,m}, s_{u,m}, q_{u,m}, d_s, d_g, t_c, t_s, t_g, t_r\}$  are known numbers that can be found in Table 2,  $\{\mu_{LD}, \mu_{LL}\}$  are prescribed load mean values, and  $\theta = \{B, L\}$  are the design parameters.

### 3. Reliability analysis

Let us now consider the following drilled shaft constructed at the site:  $B = 1.2$  m,  $L = 74.7$  m, and also let  $\mu_{LD} = 8000$  kN and  $\mu_{LL} = 4000$  kN. Four reliability methods will be presented in this chapter to determine the reliability of this particular shaft, including (a) Direct Monte Carlo simulation (MCS) (Ang & Tang, 1984); (b) first-order reliability method (FORM) (Hasofer & Lind, 1974; Der Kiureghian, 2000; Liu and Der Kiureghian, 1991); (c) importance sampling (IS) (Melchers, 1989; Hohenbichler & Rackwitz, 1988; Der Kiureghian & Dakessian, 1998; Au et al., 1999); and (d) subset simulation (Subsim) (Au & Beck, 2001). Note that only these four methods are reviewed due to their simplicity and practicality. More sophisticated methods are not the main theme of this chapter. The FORM is not a Monte Carlo based method. It is presented here because the IS method requires the FORM solution.

By definition, reliability is unity minus failure probability. As a result, central to reliability analysis is the determination of the failure probability, the probability that the failure event  $F$  occurs, denoted by  $P(F|\theta)$ . The failure probability can be found by evaluating the following integral:

$$P(F|\theta) = \int 1(SR(z,\theta) < 1)p(z|\theta) dz \tag{9}$$

where  $1(\cdot)$  is the indicator function: it is unity if the statement is true and is zero otherwise. When the  $Z$  dimension ( $p$ ) is high, the numerical solution for this integral is typically infeasible. A possible remedy is to adopt the Monte Carlo simulation to evaluate this integral.

#### 3.1 Direct Monte Carlo simulation

According to the Law of Large Number (Ang and Tang, 1984), the integral can be approximately evaluated as follows:

$$P(F|\theta) \approx \frac{1}{N} \sum_{i=1}^N 1(SR(Z^{(i)}, \theta) < 1) \equiv P_F^{MCS} \tag{10}$$

where  $N$  is the total number of MCS independent samples;  $Z^{(i)}$  is the  $i$ -th sample of  $Z$ , drawn from the jointly standard Gaussian distribution  $p(z|\theta)$ ;  $P_F^{MCS}$  is the estimator for  $P(F|\theta)$  based on MCS. This estimator is unbiased, i.e., the expected value of  $P_F^{MCS}$  is exactly  $P(F|\theta)$ , and is with c.o.v. =  $\{[1 - P(F|\theta)]/N/P(F|\theta)\}^{0.5}$ . Note that the c.o.v. does not depend on  $Z$  dimension and does not depend on the complexity of the problem, either. This is the key advantage for MCS, especially for geotechnical problems where nonlinearity and uncertainty dimension is usually high.

The disadvantage of MCS is that it may require a large sample size when  $P(F|\theta)$  is small. A rule of thumb is that it requires  $N = 10/P(F|\theta)$  to achieve a reasonable accuracy, i.e., c.o.v. = 30%. The disadvantage is acceptable when the calculation of  $SR$  is fast, e.g., the example design problem for drilled shaft. For problems where a single calculation of  $SR$  is time consuming, e.g., finite element analysis, MCS may be infeasible.

For the example design problem of a drilled shaft, the following steps can be taken to estimate  $P(F | \theta)$ :

- a. Draw  $N$  independent samples of  $Z^{(i)} = \{Z^{(i)}_{\sigma'vsr}, Z^{(i)}_{\sigma'vgr}, Z^{(i)}_{su}, Z^{(i)}_{qu}, Z^{(i)}_{Sc}, Z^{(i)}_{Ssr}, Z^{(i)}_{Sgr}, Z^{(i)}_{Srr}, Z^{(i)}_{LD}, Z^{(i)}_{LL}\}$  from the jointly standard Gaussian distribution.
- b. For each sample set, evaluate

$$SR(Z^{(i)}, \theta) = \frac{\pi B \left[ e^{2.7+0.3[\ln(70)-0.2z_{su}^{(i)}]+0.3216z_{Sc}^{(i)}} \cdot 42.1 + e^{1.0802-0.6588\ln(45.85)+[\ln(350)-0.1z_{\sigma'rs}^{(i)}]+0.5414z_{Sg}^{(i)}} \cdot 7.5 + e^{2.1792-0.7528\ln(59.65)+[\ln(480)-0.1z_{\sigma'vg}^{(i)}]+0.6689z_{Sg}^{(i)}} \cdot 20.1 + e^{3.0253+0.414[\ln(900)-0.5z_{qu}^{(i)}]+0.7160z_{Srr}^{(i)}} \cdot 5.0 \right]}{e^{\ln(8000/\sqrt{1+0.1^2})+\sqrt{\ln(1+0.1^2)} \cdot z_{LD}^{(i)}} + e^{\ln(4000/\sqrt{1+0.25^2})+\sqrt{\ln(1+0.25^2)} \cdot z_{LL}^{(i)}}} \quad (11)$$

c. Let

$$P_F^{MCS} = \frac{1}{N} \sum_{i=1}^N 1(SR(Z^{(i)}, \theta) < 1) \quad (12)$$

By taking  $N = 10^6$ ,  $P_F^{MCS}$  is found to be around  $5.9 \times 10^{-4}$ . Note that  $P_F^{MCS}$  is in general not the same as the actual  $P(F | \theta)$  but is only its estimator. It will be informative to also know the c.o.v. of  $P_F^{MCS}$ . This c.o.v. can be estimated as  $\{[1 - P_F^{MCS}]/N/P_F^{MCS}\}^{0.5} = 4\%$ . Figure 1 shows a conceptual plot for the MCS samples. These samples center at the origin, the location of the mean value of  $Z$ . Since the failure probability is small, most samples are in the non-failure region ( $SR > 1$ ), while only few samples are in the failure region ( $SR < 1$ ). Since the c.o.v. of  $P_F^{MCS}$  is  $\{[1 - P(F | \theta)]/N/P(F | \theta)\}^{0.5} \approx [1/(\# \text{ of failure samples})]^{0.5}$ , the disadvantage of MCS is due to lacking of failure samples.

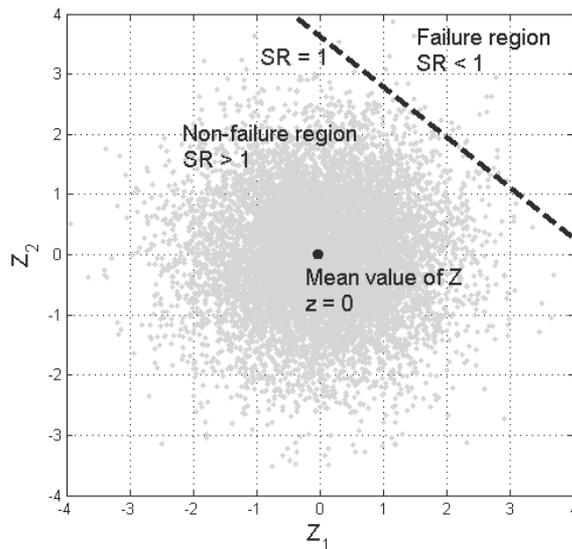


Fig. 1. Conceptual plot for the Monte Carlo samples in the standard Gaussian space

### 3.2 First-order reliability method

The first-order reliability method (FORM) (Hasofer & Lind, 1974; Der Kiureghian, 2000; Liu and Der Kiureghian, 1991) is not a Monte-Carlo based method. It is introduced herein because it is the most popular reliability method that is used in civil engineering problems and also because the forthcoming Monte-Carlo based method, importance sampling, requires the knowledge of FORM. FORM is based on a mathematical fact that the shortest distance between the limit-state line  $SR(z, \theta) = 1$  to the origin  $z = 0_{p \times 1}$  is closely related to  $P(F | \theta)$  (see Figure 2). In fact, this distance is called the reliability index  $\beta$ , and it can be shown that  $P(F | \theta)$  is roughly equal to  $\Phi(-\beta)$  for relatively simple problems, where  $\Phi$  is the cumulative density function of standard Gaussian distribution. As a result, determining  $P(F | \theta)$  is equivalent to finding the shortest distance by the following optimization problem:

$$\min_z \|z\| \quad \text{subjected to} \quad SR(z, \theta) = 1 \quad (13)$$

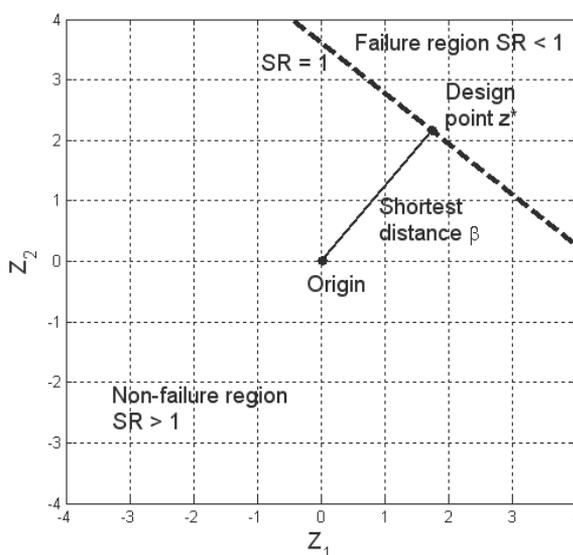


Fig. 2. Significance of the design point in the standard Gaussian space

The solution point of Eq. (13) is called the design point  $z^*$ . For problems with differentiable SR, the following necessary conditions hold for the design point: (a)  $SR(z^*, \theta) = 1$  and (b) the gradient vector of SR at  $z^*$ , i.e.,  $\nabla_z SR(z^*, \theta)$ , is parallel to  $z^*$ . The reliability index is simply the length of  $z^*$ . There are many algorithms for finding the design point  $z^*$ , but the following one is among the simplest (Ang and Tang, 1984):

- a. Initialize  $z_0^*$  at any location
- b. Evaluate  $\nabla_z SR(z_0^*, \theta) = \left[ \frac{\partial SR(z_0^*, \theta)}{\partial z_1} \quad \frac{\partial SR(z_0^*, \theta)}{\partial z_2} \quad \dots \quad \frac{\partial SR(z_0^*, \theta)}{\partial z_p} \right]$ . This may require numerical approximations for the partial derivatives.
- c. Find  $\alpha_0$  such that  $SR\left(\alpha_0 \cdot \nabla_z SR(z_0^*, \theta)^T, \theta\right) = 1$  and let  $z_1^* = \alpha_0 \cdot \nabla_z SR(z_0^*, \theta)^T$ . It may require a Newton-method search for determining  $\alpha_0$ .

Cycle the Steps b-c until convergence. Once the algorithm converges, it is clear that the converging solution satisfy the two necessary conditions at the same time. For problems with very small failure probability, FORM can be much more efficient than MCS because the former does not require as many SR calculations as the latter. However, for problems with high dimensional  $Z$ , the optimization problem in Eq. (13) may become extremely challenging and even become intractable.

For the example design problem of a drilled shaft, the gradient vector is simply

$$\begin{aligned} \nabla_z SR(z, \theta) &= \left[ \frac{\partial SR(z, \theta)}{\partial z_{su}} \quad \frac{\partial SR(z, \theta)}{\partial z_{vs}} \quad \dots \right] \\ &= \pi B \left[ -0.6 \times 42.1 \times e^{2.7+0.3[\ln(70)-0.2z_{su}]+0.3216z_{vs}} \quad -0.1 \times 7.5 \times e^{1.0802-0.6588\ln(45.85)+[\ln(350)-0.1z_{vs}]+0.5414z_{ss}} \quad \dots \right] \end{aligned} \quad (14)$$

The above steps are taken to find the design point  $z^*$ , which is found to satisfy  $z_{su}^* = -0.245$ ,  $z_{sc}^* = -1.333$ ,  $z_{vs}^* = -0.121$ ,  $z_{ss}^* = -0.664$ ,  $z_{vsg}^* = -0.290$ ,  $z_{sg}^* = -1.953$ ,  $z_{qu}^* = -0.345$ ,  $z_{sr}^* = -1.268$ ,  $z_{LD}^* = 0.644$ , and  $z_{LL}^* = 0.894$ . Note that for all stabilizing variables, the design point coordinates are negative, and for the two destabilizing variables  $L_D$  and  $L_L$ , the design point coordinates are positive. The distance from the design point to the origin is shortest distance is 3.02, so the estimated  $P(F|\theta)$  is equal to  $\Phi(-3.02) = 1.3 \times 10^{-3}$ . This result is an approximation to the actual value of  $P(F|\theta)$ .

### 3.3 Importance sampling

As mentioned before, for problems with small failure probability, the disadvantage of MCS is that it may require many samples to obtain sufficient failure samples. The importance sampling (IS) (Melchers, 1989; Hohenbichler & Rackwitz, 1988; Der Kureghian & Dakessian, 1998; Au et al., 1999) method mitigates this issue by shifting the standard Gaussian distribution  $p(z|\theta)$  to a new center that is closer to the failure region. The most logical choice of this new center is the design point  $z^*$  from FORM. Let the shifted distribution be  $q(z|\theta)$ :

$$q(z|\theta) = \frac{1}{2\pi^{p/2}} e^{-\frac{1}{2}(z-z^*)^T(z-z^*)} \quad (15)$$

It is clear that

$$\begin{aligned} P(F|\theta) &= \int \mathbf{1}(SR(z, \theta) < 1) \frac{p(z|\theta)}{q(z|\theta)} q(z|\theta) dz \\ &= \int \mathbf{1}(SR(z, \theta) < 1) \cdot e^{-\frac{1}{2}z^T z + \frac{1}{2}(z-z^*)^T(z-z^*)} q(z|\theta) dz \end{aligned} \quad (16)$$

According to the Law of Large Number,

$$P(F|\theta) \approx \frac{1}{N} \sum_{i=1}^N \mathbf{1}(SR(Z^{(i)}, \theta) < 1) \cdot e^{\frac{1}{2}z^{*T} z^* - z^{*T} Z^{(i)}} \equiv P_F^{IS} \quad (17)$$

where the samples  $Z^{(i)}$  are drawn from the shifted distribution  $q(z|\theta)$ . Figure 3 shows the conceptual plots for the samples from the IS method: roughly one half of the samples falling

into the failure region. As a result, the c.o.v. for the IS estimator  $P_F^{IS}$  can be much smaller than that for  $P_F^{MCS}$ .

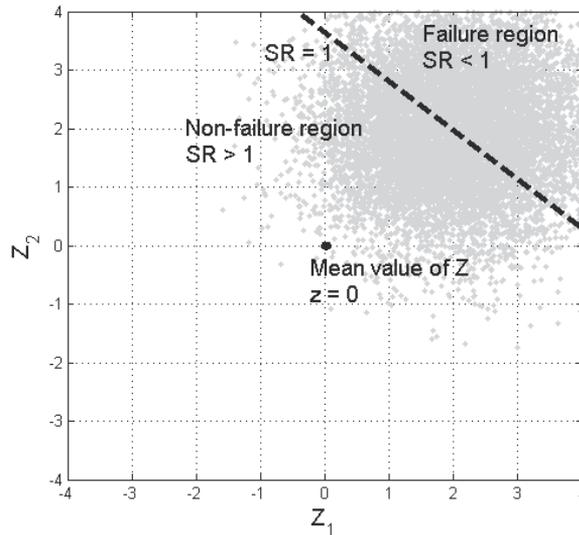


Fig. 3. Conceptual plot for the IS samples in the standard Gaussian space

For the example design problem of a drilled shaft, the following steps can be taken to estimate  $P(F | \theta)$ :

- Find the design point  $z^*$  for FORM.
- Draw  $N$  independent samples of  $Z^{(i)} = \{Z^{(i)}_{\sigma'_{vs}}, Z^{(i)}_{\sigma'_{vg}}, Z^{(i)}_{su}, Z^{(i)}_{qu}, Z^{(i)}_{Sc}, Z^{(i)}_{Ssr}, Z^{(i)}_{Sgr}, Z^{(i)}_{Sr}, z^{(i)}_{LD}, z^{(i)}_{LL}\}$  from the shifted distribution  $q(z | \theta)$ .
- For each sample set, evaluate

$$SR(Z^{(i)}, \theta) = \frac{\pi B \left[ e^{2.7+0.3[\ln(70)-0.2z_{qu}^{(i)}]+0.3216z_{Sc}^{(i)}} \cdot 42.1 + e^{1.0802-0.6588\ln(45.85)+[\ln(350)-0.1z_{\sigma'_{vs}}^{(i)}]+0.5414z_{Ss}^{(i)}} \cdot 7.5 + e^{2.1792-0.7528\ln(59.65)+[\ln(480)-0.1z_{\sigma'_{vg}}^{(i)}]+0.6689z_{Sg}^{(i)}} \cdot 20.1 + e^{3.0253+0.414[\ln(900)-0.5z_{qu}^{(i)}]+0.7160z_{Sr}^{(i)}} \cdot 5.0 \right]}{e^{\ln(8000/\sqrt{1+0.1^2})+\sqrt{\ln(1+0.1^2)} \cdot z_{LD}^{(i)}} + e^{\ln(4000/\sqrt{1+0.25^2})+\sqrt{\ln(1+0.25^2)} \cdot z_{LL}^{(i)}}} \quad (18)$$

- Let

$$P_F^{IS} = \frac{1}{N} \sum_{i=1}^N 1(SR(Z^{(i)}, \theta) < 1) \cdot e^{\frac{1}{2}z^{*T} z^* - z^{*T} Z^{(i)}} \quad (19)$$

By taking  $N = 1000$ ,  $P_F^{IS}$  is found to be around  $6.1 \times 10^{-4}$ . Its c.o.v. is estimated to be around 7%. Compared to MCS using  $10^6$  samples yielding a 4% c.o.v., the IS method is much more efficient. It seems like the IS method improves MCS, but in fact this is not entirely true: as reported in Au & Beck (2003), the IS method may suffer from the issue of high  $Z$  dimension (as FORM does), but MCS does not have such limitation.

### 3.4 Subset simulation

Among the previous reliability methods, no method is suitable for complex problems with dimensional  $Z$  and with small failure probability. FORM is suitable for problems with small failure probability but not for those with high dimensional  $Z$ . In contrast, MCS is robust with  $Z$  dimension and problem complexity but may be inefficient for problems with small failure probability. The IS method also suffers from the issue of high dimensional  $Z$ .

Subset simulation (Subsim) (Au & Beck, 2001) is among the few reliability methods that are robust against all the aforementioned aspects. Subsim inherits most advantages of MCS: it is robust against  $Z$  dimension and problem complexity, but its computational cost for problems with small failure probability is typically acceptable. The basic idea of Subsim is to express the failure probability  $P(F|\theta)$  as a product of several larger conditional probabilities, so that the estimation of  $P(F|\theta)$  can be achieved by estimating the conditional probabilities and multiply them together.

Let us first introduce intermediate failure events  $\{F_1, F_2, \dots, F_m\}$ . For our purpose, these failure events can be defined as

$$F_i = \{SR(z, \theta) < b_i\} \quad i = 1, \dots, m \quad (20)$$

where  $b_1 > b_2 > \dots > b_m = 1$ . It is then clear that the intermediate failure events are nested, i.e.,  $F_1 \supset F_2 \supset \dots \supset F_m = F$ . Moreover, the failure event  $F$  is the intersection of all intermediate failure events. According to the operation of conditional probability,

$$\begin{aligned} P(F|\theta) &= P\left(\bigcap_{i=1}^m F_i|\theta\right) = P\left(F_m \mid \bigcap_{i=1}^{m-1} F_i, \theta\right) \cdots P(F_2|F_1, \theta) \cdot P(F_1|\theta) \\ &= P(F_m|F_{m-1}, \theta) \cdots P(F_2|F_1, \theta) \cdot P(F_1|\theta) = \prod_{i=2}^m P(F_i|F_{i-1}, \theta) \cdot P(F_1|\theta) \end{aligned} \quad (21)$$

As a result, the estimation of  $P(F|\theta)$  can be achieved through the estimation of the conditional probabilities  $P(F_1|\theta)$ ,  $P(F_2|F_1, \theta)$ ,  $\dots$ ,  $P(F_m|F_{m-1}, \theta)$ . Note that although  $P(F|\theta)$  may be very small, the conditional probabilities  $P(F_1|\theta)$ ,  $\dots$ ,  $P(F_m|F_{m-1}, \theta)$  can be made large and can be estimated in a more accurate manner. Hence, the issue of small failure probability for MCS is resolved. In the following, the estimation of these conditional probabilities will be addressed.

#### Estimation of $P(F_1|\theta)$

This estimation can be easily done by using MCS, i.e., draw  $N_0$  samples of  $Z$  from the standard Gaussian distribution, denoted by  $\{Z_0^{(k)}: k=1, \dots, N_0\}$ . Then,

$$P(F_1|\theta) \approx \frac{1}{N_0} \sum_{i=1}^{N_0} 1(SR(Z_0^{(k)}) < b_1) \equiv P_1^{SS} \quad (22)$$

Note that  $P(F_1|\theta)$  is typically quite large, hence the c.o.v. of  $P_1^{SS}$  is typically small. Among the  $N_0$  samples, let there be  $R_0$  samples, denoted by  $\{Z_0^{*(k)}: k=1, \dots, R_0\}$ , satisfying  $SR < b_1$ . Let us call these samples the below- $b_1$  samples. These samples are actually distributed as  $p(z|F_1, \theta)$ , which can be expressed as

$$p(z | F_1, \theta) = \frac{p(z | \theta) \cdot 1(SR(z, \theta) < b_1)}{P(F_1 | \theta)} = \frac{e^{-\frac{1}{2}z^T z} \cdot 1(SR(z, \theta) < b_1)}{(2\pi)^{p/2} P(F_1 | \theta)} \quad (23)$$

**Estimation of P(F<sub>2</sub> | F<sub>1</sub>, θ)**

Suppose we know how to draw N<sub>1</sub> samples, denoted by {Z<sub>1</sub><sup>(k)</sup>: k=1, ..., N<sub>1</sub>}, from p(z | F<sub>1</sub>, θ). The estimation of P(F<sub>2</sub> | F<sub>1</sub>, θ) can then be made:

$$P(F_2 | F_1, \theta) \approx \frac{1}{N_1} \sum_{k=1}^{N_1} 1(SR(Z_1^{(k)}, \theta) < b_2) \equiv P_2^{SS} \quad (24)$$

However, drawing samples from p(z | F<sub>1</sub>, θ) is nontrivial. Recall that the below-b<sub>1</sub> samples {Z<sub>0</sub><sup>\*(k)</sup>: k=1, ..., R<sub>0</sub>} from stage 1 are already distributed as p(z | F<sub>1</sub>, θ). It is then possible to use the Metropolis algorithm (Au & Beck, 2001) to generate more samples that are also distributed as p(z | F<sub>1</sub>, θ). Each below-b<sub>1</sub> sample Z<sub>0</sub><sup>\*(k)</sup> is taken to be the initial sample of a Markov chain whose stationary distribution is p(z | F<sub>1</sub>, θ). Let further the j-th sample of the k-th Markov chain be Z<sub>1</sub><sup>(k,j)</sup>. The k-th below-b<sub>1</sub> sample Z<sub>0</sub><sup>\*(k)</sup> from stage 1 is therefore Z<sub>1</sub><sup>(k,1)</sup>. The following Metropolis algorithm can then be taken to generate the rest samples {Z<sub>1</sub><sup>(k,j)</sup>: j = 2, ...} for the k-th Markov chain:

- a. Given the j-th sample Z<sub>1</sub><sup>(k,j)</sup> in this chain, draw a candidate sample Z<sub>1</sub><sup>C</sup> from a Gaussian distribution centered at Z<sub>1</sub><sup>(k,j)</sup> and with a chosen covariance matrix Σ.
- b. Compute the following ratio r:

$$r = \frac{p(Z_1^C | F_1, \theta)}{p(Z_1^{(k,j)} | F_1, \theta)} = e^{-\frac{1}{2}Z_1^C T Z_1^C + \frac{1}{2}Z_1^{(k,j)T} Z_1^{(k,j)}} \cdot \frac{1[SR(Z_1^C, \theta) < b_1]}{1[SR(Z_1^{(k,j)}, \theta) < b_1]} \quad (25)$$

- c. Accept the candidate sample, i.e., let Z<sub>1</sub><sup>(k,j+1)</sup> be Z<sub>1</sub><sup>C</sup>, with probability min(1,r), and repeat the previous sample, i.e., let Z<sub>1</sub><sup>(k,j+1)</sup> be Z<sub>1</sub><sup>(k,j)</sup>, with probability 1-min(1,r).

Suppose each Markov chain generate M<sub>1</sub> samples, the following samples {Z<sub>1</sub><sup>(k,j)</sup>: k = 1, ..., R<sub>0</sub>, j = 1, ..., M<sub>1</sub>} are available. All these samples are distributed as p(z | F<sub>1</sub>, θ), so there are N<sub>1</sub> = R<sub>0</sub>M<sub>1</sub> samples, rearranged to be {Z<sub>1</sub><sup>(k)</sup>: k = 1, ..., N<sub>1</sub>}, for the estimation of P(F<sub>2</sub> | F<sub>1</sub>, θ) in Eq. (24). Among these N<sub>1</sub> samples, let there be R<sub>1</sub> samples, denoted by {Z<sub>1</sub><sup>\*(i)</sup>: i=1, ..., R<sub>1</sub>}, satisfying SR < b<sub>2</sub>. These samples are the below-b<sub>2</sub> samples and are actually distributed as p(z | F<sub>2</sub>, θ), and the same Metropolis algorithm can be used to generate more samples from p(z | F<sub>2</sub>, θ) to estimate P(F<sub>3</sub> | F<sub>2</sub>, θ). This process continues until P(F<sub>m</sub> | F<sub>m-1</sub>, θ) is estimated. Finally, P(F | θ) can be estimated as

$$P(F | \theta) \approx \prod_{i=1}^m P_i^{SS} = \prod_{i=1}^m \frac{R_{i-1}}{N_{i-1}} \equiv P_F^{SS} \quad (26)$$

In real application of Subsim, the threshold b<sub>i</sub> is adaptively chosen so that R<sub>i-1</sub> = N<sub>i-1</sub>/10 (except the final stage R<sub>m-1</sub> is not equal to N<sub>m-1</sub>/10), i.e., b<sub>i</sub> is taken to be the 10% percentile of {SR(Z<sub>i-1</sub><sup>(k)</sup>, θ): k = 1, ..., N<sub>i-1</sub>}. Moreover, each Markov chain generate 10 samples, including the below-b<sub>i-1</sub> sample from the previous stage. This makes N<sub>0</sub> = N<sub>1</sub> = ... = N<sub>m-1</sub> = N<sub>SS</sub>. Au and Beck (2001) show that the estimator P<sub>F</sub><sup>SS</sup> is asymptotically unbiased. Ching et al. (2005) show that the c.o.v. of P<sub>F</sub><sup>SS</sup> [δ(P<sub>F</sub><sup>SS</sup>)] is bounded by

$$\sqrt{\frac{1}{N_{SS}} \left( 9(m-1) + \frac{1 - P_m^{SS}}{P_m^{SS}} \right)} \leq \delta(P_F^{SS}) \leq \sqrt{\frac{1}{N_{SS}} \left( 9 + 99(m-2) + \frac{1 - P_m^{SS}}{P_m^{SS}} \cdot 11 \right)} \quad (27)$$

For the example design problem of a drilled shaft, 1000 MCS samples of  $Z$  ( $N_{SS} = 1000$ ) are drawn from the standard Gaussian distribution for the first stage. For each sample  $Z^{(i)}$ , its SR sample value is evaluated to get  $SR(Z^{(i)}, \theta)$ . The leftmost plot in Figure 4 shows such SR samples. It is clear that there are no failure samples, i.e., samples satisfying  $SR < 1$ . The first threshold  $SR = b_1$ , shown in the left plot as the horizontal line, is then identified as the 10% percentile of the SR sample values, and  $F_1$  event is therefore  $\{SR(z, \theta) < b_1\}$ . As a result,  $P(F_1 | \theta) \approx P_1^{SS} = 0.1$ . The 100 below- $b_1$  samples (the darker dots) are distributed as  $p(z | F_1, \theta)$ . These below- $b_1$  samples are then taken in the Metropolis algorithm to generate more samples also distributed as  $p(z | F_1, \theta)$ : each below- $b_1$  sample is taken to lead a Markov chain that generated 9 more samples distributed as  $p(z | F_1, \theta)$ . These 1000 new samples are seen in the middle left plot. Note that all these samples have SR values less than  $b_1$  because they are distributed as  $p(z | F_1, \theta)$ .

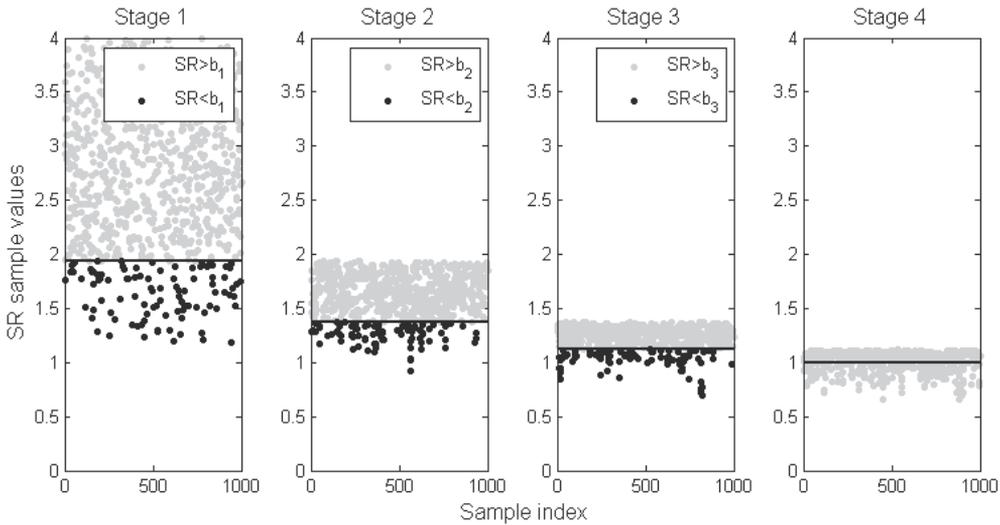


Fig. 4. Evolution of the SR samples in various stages (stage 1 to 4 from left to right) for Subsim

The second threshold  $SR = b_2$ , shown in the middle left plot as the horizontal line, is then identified as the 10% percentile of the SR sample values, and  $F_2$  event is therefore  $\{SR(Z, \theta) < b_2\}$ . As a result,  $P(F_2 | F_1, \theta) \approx P_2^{SS} = 0.1$ . Similarly, the 100 below- $b_2$  samples (the darker dots) are distributed as  $p(z | F_2, \theta)$ . These below- $b_2$  samples are then taken in the Metropolis algorithm to generate 1000 samples also distributed as  $p(z | F_2, \theta)$ , i.e., the samples seen in the rightmost plot.

The third stage is similar to the previous stages (see the middle right plot). Similarly, the threshold  $b_3$  is adaptively chosen, and  $P(F_3 | F_2, \theta) \approx P_3^{SS} = 0.1$ . The fourth stage is somewhat different because now the SR values of the 1000 samples distributed as  $p(z | F_3, \theta)$  (the gray dots in the rightmost plot) are close to the failure threshold  $b = 1$ . The 10% percentile of the

SR values is found to be less than 1, i.e.,  $P(F | F_3, \theta) > 0.1$ . In fact, 28.9% of 1000 SR values are less than 1. In this scenario, the fourth threshold  $b_4$  is no longer adaptively chosen as the 10% percentile but is taken to be 1, and the entire Subsim algorithm ends at this stage. Consequently,  $P(F_4 | F_3, \theta) = P(F | F_3, \theta) \approx P_4^{SS} = 0.289$ , and the Subsim estimate for  $P(F | \theta)$  is simply  $P_F^{SS} = P_1^{SS} \times P_2^{SS} \times P_3^{SS} \times P_4^{SS} = 2.89 \times 10^{-4}$ . The bounds for this estimator can be found to be

$$\sqrt{\frac{1}{1000} \left( 9 \times 3 + \frac{1 - 0.289}{0.289} \right)} = 17.2\% \leq \delta(P_F^{SS}) \leq 48.4\% = \sqrt{\frac{1}{1000} \left( 9 + 99 \times 2 + \frac{1 - 0.289}{0.289} \cdot 11 \right)} \quad (28)$$

#### 4. Reliability-based design

In this section, state-of-the-art developments in the Monte Carlo based reliability-based design methods are reviewed. Reliability-based design (RBD) is the inverse problem of reliability analysis: the purpose of reliability analysis is to obtain the reliability given the design dimension of an engineering system, but the purpose of RBD is to design for the dimension that provides the target reliability. The Monte Carlo based methods recently developed by Ching & Phoon (2010) will be introduced. These Monte Carlo based methods are able to convert the RBD design constraint into simple algebraic design equations. Moreover, these methods inherit most of the advantages of MCS, i.e., not limited by the Z dimension, problem complexity, etc. One limitation is that the number of design parameters, e.g., dimension of  $\theta$ , cannot be too large, which is usually the case for geotechnical designs. For RBD, the objective is to enforce the following *probabilistic* constraint during the design process:

$$P(SR(Z, \theta) < 1 | \theta) = \int p(z | \theta) \cdot 1(SR(z, \theta) < 1) dz \leq P_F^* \quad (29)$$

where  $P_F^*$  is the target failure probability;  $1(\cdot)$  is the indicator function, i.e. it is equal to 1 if the argument (safety ratio less than one) is true; otherwise, it is zero. The purpose of this section is to show that this probabilistic design constraint can be transformed into a deterministic *algebraic* constraint of the following format:

$$c(\theta) \geq 1 \quad (30)$$

Intuitively,  $c(\theta)$  should be taken to be a conservative version of  $SR(z^*, \theta)$ , i.e.,  $c(\theta) < SR(z^*, \theta)$ , where  $z^*$  is the characteristic value of Z, which can be taken to be the mean value or median value of Z. As a result, requiring  $c(\theta) \geq 1$  is much stronger than requiring  $SR(z^*, \theta) \geq 1$ , leading to a more conservative design with a small target failure probability  $P_F^*$ . Depending on how this conservatism is applied, there are four possible algebraic design formats:

a. Safety factor design

For this design format,  $c(\theta)$  is taken to be the  $\eta$  quantile ( $\eta$  is called the probability threshold) of  $SR(Z, \theta)$ , denoted by  $SR^\eta(\theta)$ . The probability threshold  $\eta$  is typically taken to be a small number, e.g., 0.05, to ensure a conservative design. For a normally distributed SR, the 5% quantile,  $SR^{0.05} = \mu_{SR} (1 - 1.645 \delta_{SR})$ , in which  $\mu_{SR}$  and  $\delta_{SR}$  are the mean and c.o.v. of SR. This definition is sensible because we are applying a value less than the mean. Requiring  $c(\theta) \geq 1$  is equivalent to requiring

$$c(\theta) = \frac{SR(z^*, \theta)}{SR(z^*, \theta)/c(\theta)} = \frac{SR(z^*, \theta)}{SR(z^*, \theta)/SR^\eta(\theta)} = \frac{SR(z^*, \theta)}{SF^\eta} \geq 1 \quad (31)$$

where SF is the safety factor, which clearly depends on  $\eta$ . Note that the required SF is simply the nominal safety ratio  $SR(z^*, \theta)$  divided by the  $\eta$  quantile  $SR^\eta(\theta)$ .

b. Load-resistance factor design (LRFD)

For many geotechnical design problems, the safety ratio  $SR(z, \theta)$  has the format of

$$SR(z, \theta) = \frac{S(z, \theta)}{L_D(z, \theta) + L_L(z, \theta)} \quad (32)$$

where S is the total resistance, and  $L_D$  and  $L_L$  are the dead and live loads. In the case, a possible choice for  $c(\theta)$  is

$$c(\theta) = \frac{S^\eta(\theta)}{L_D^{1-\eta}(\theta) + L_L^{1-\eta}(\theta)} \quad (33)$$

where  $S^\eta(\theta)$  is the  $\eta$  quantile of  $S(Z, \theta)$ , and  $L_D^{1-\eta}(\theta)$  and  $L_L^{1-\eta}(\theta)$  are the  $1-\eta$  quantiles of  $L_D(Z, \theta)$  and  $L_L(Z, \theta)$ . The same probability threshold  $\eta$  (e.g., 0.05) is applied to all three random variables. This would not only ensure a conservative design but also ensure that all random variables have the same exceedance/non-exceedance probability over the corresponding quantiles. Requiring  $c(\theta) \geq 1$  is equivalent to requiring

$$c(\theta) = \frac{\frac{S^\eta(\theta)}{S(z^*, \theta)} S(z^*, \theta)}{\frac{L_D^{1-\eta}(\theta)}{L_D(z^*, \theta)} L_D(z^*, \theta) + \frac{L_L^{1-\eta}(\theta)}{L_L(z^*, \theta)} L_L(z^*, \theta)} = \frac{\gamma_S^\eta \cdot S(z^*, \theta)}{\gamma_{LD}^\eta \cdot L_D(z^*, \theta) + \gamma_{LL}^\eta \cdot L_L(z^*, \theta)} \geq 1 \quad (34)$$

where  $\gamma_S$  is the resistance factor, while  $\gamma_{LD}$  and  $\gamma_{LL}$  are the load factors. These factors clearly depend on  $\eta$ . It is also clear that  $\gamma_S < 1$  and that  $\gamma_{LD}, \gamma_{LL} > 1$  if the probability threshold  $\eta$  is small. Note that the required resistance factor is simply the  $\eta$  quantile  $S^\eta(\theta)$  divided by the nominal resistance  $S(z^*, \theta)$ , and the required load factor is simply the  $1-\eta$  quantile  $L^{1-\eta}(\theta)$  divided by the nominal load  $L(z^*, \theta)$ .

c. Multiple resistance factor design (MRFD)

For some geotechnical design problems, the total resistance S is contributed by several different components. Let us denote the various components by  $S_x$ . For the drilled shaft example, the subscript 'x' can be either 'c', 's', 'g', or 'r', depending on which stratum provides the side resistance, and

$$SR(z, \theta) = \frac{S_c(z, \theta) + S_s(z, \theta) + S_g(z, \theta) + S_r(z, \theta)}{L_D(z, \theta) + L_L(z, \theta)} \quad (35)$$

In the case, a possible choice for  $c(\theta)$  is

$$c(\theta) = \frac{S_c^\eta(\theta) + S_s^\eta(\theta) + S_g^\eta(\theta) + S_r^\eta(\theta)}{L_D^{1-\eta}(\theta) + L_L^{1-\eta}(\theta)} \quad (36)$$

Again, the same probability threshold  $\eta$  is applied to all six random variables to ensure that all random variables have the same exceedance/non-exceedance probability over the corresponding quantiles. Requiring  $c(\theta) \geq 1$  is equivalent to requiring

$$c(\theta) = \frac{\frac{S_c^\eta(\theta)}{S_c(z^*, \theta)} S_c(z^*, \theta) + \dots + \frac{S_r^\eta(\theta)}{S_r(z^*, \theta)} S_r(z^*, \theta)}{\frac{L_D^{1-\eta}(\theta)}{L_D(z^*, \theta)} L_D(z^*, \theta) + \frac{L_L^{1-\eta}(\theta)}{L_L(z^*, \theta)} L_L(z^*, \theta)} \quad (37)$$

$$= \frac{\gamma_{S_c}^\eta \cdot S_c(z^*, \theta) + \dots + \gamma_{S_r}^\eta \cdot S_r(z^*, \theta)}{\gamma_{L_D}^\eta \cdot L_D(z^*, \theta) + \gamma_{L_L}^\eta \cdot L_L(z^*, \theta)} \geq 1$$

where  $\gamma_{S_c}$ ,  $\gamma_{S_s}$ ,  $\gamma_{S_g}$ , and  $\gamma_{S_r}$  are the resistance factors, while  $\gamma_{L_D}$  and  $\gamma_{L_L}$  are the load factors. It is clear that all resistance factors are less than 1 and that  $\gamma_{L_D}, \gamma_{L_L} > 1$  if the probability threshold  $\eta$  is small. Note that the required resistance factor is simply the  $\eta$  quantile  $S^\eta(\theta)$  divided by the nominal resistance  $S(z^*, \theta)$ , and the required load factor is simply the  $1-\eta$  quantile  $L^{1-\eta}(\theta)$  divided by the nominal load  $L(z^*, \theta)$ .

d. Partial factor design

In contrast to LRFD and MRFD where the factors are applied to load and resistance terms, the partial-factor design format applies the (partial) factors to  $Z$  directly, i.e.,

$$c(\theta) = SR(z_1^\eta, z_2^\eta, \dots, z_p^{1-\eta}, \theta) \geq 1 \quad (38)$$

where either the  $\eta$  or  $1-\eta$  quantile of  $Z_i$  is taken depending on its characteristic. For  $Z_i$  that is clearly stabilizing,  $\eta$  quantile of  $Z_i$  should be taken, while  $1-\eta$  quantile should be adopted for destabilizing  $Z_i$ . For  $Z_i$  that is not influential or whose effect cannot be clearly discerned as stabilizing or destabilizing, the mean or median value may be taken. Requiring  $c(\theta) \geq 1$  is equivalent to requiring

$$c(\theta) = SR\left(\frac{z_1^\eta}{z_1^*}, \frac{z_2^\eta}{z_2^*}, \dots, \frac{z_p^{1-\eta}}{z_p^*}, \theta\right) = SR\left(\gamma_1^\eta z_1^*, \gamma_2^\eta z_2^*, \dots, \gamma_p^\eta z_p^*, \theta\right) \geq 1 \quad (39)$$

where  $\gamma$ 's are the partial factors. For stabilizing random variables, the partial factors are less than 1, and for destabilizing random variables, the partial factors are greater than 1. Note that the required partial factor is simply the  $\eta$  quantile  $z_i^\eta$  divided by its characteristic value  $z_i^*$  if  $z_i$  is stabilizing and is the  $1-\eta$  quantile  $z_i^{1-\eta}$  divided by its characteristic value  $z_i^*$  if  $z_i$  is destabilizing.

Note that algebraic design constraints based on the above four formats are convenient, because to make sure  $c(\theta) \geq 1$  holds, it only requires a *single* algebraic evaluation of  $SR(Z, \theta)$  using deterministic factors  $\gamma$ 's and characteristic values  $z^*$ . However, RBD is more theoretically involved and computationally demanding: in order to verify if Eq. (29) holds, it requires a reliability analysis, which in the most general case, may take millions of Monte Carlo evaluations of  $SR(Z, \theta)$ . If the equivalence between the four algebraic design formats and the RBD can be established, it will be significant in the following practical sense:

- a. One can then achieve a RBD by using any of the four algebraic formats, which is much simpler and more convenient than the former.
- b. Practical geotechnical engineers who are not familiar with reliability concept can easily achieve reliability-based design by using the equivalence.

To establish the equivalence, it is necessary to find the relation between the probability threshold  $\eta$  and the target failure probability  $P_F^*$ . It is clear that  $\eta$  controls the degree of conservatism. Recall that the algebraic design constraint is to require  $c(\theta) \geq 1$ , hence a smaller  $\eta$  will lead to a more conservative design because the value of  $c(\theta)$  decreases with decreasing  $\eta$ . As a result, it seems reasonable to use a small  $\eta$  when the target failure probability  $P_F^*$  is small, and vice versa.

#### 4.1 Statement of the equivalence principle

In Ching & Phoon (2010), it is postulated that the four algebraic design formats associated with a proper probability threshold  $\eta$  can be made equivalent to rigorous RBD based on a direct probability check. The key hypothesis needed for the principle to be practical is also clarified explicitly using a mathematical proof in Appendix. To be specific, we postulate that there exists pairs of  $(\eta, P_F^*)$  such that the following constraints are equivalent:

$$c(\theta) \geq 1 \quad (40)$$

and

$$P(SR(Z, \theta) < 1 | \theta) \leq P_F^* \quad (41)$$

Moreover, the functional relation between the pair  $(\eta, P_F^*)$  is as follows:

$$P\left(\frac{c(\theta)}{SR(Z, \theta)} > 1 | \theta\right) = P_F^* \quad (42)$$

In Eq. (42), note that the numerator is a deterministic number and the denominator is a random variable depending on  $\theta$ . Equation (42) is the key equation in the proposed algebraic design formats.

#### 4.2 Uniformity of the equivalence

The proposed approach is not practical if the relation between  $(\eta, P_F^*)$  depends on the design parameter  $\theta$ . If this happens, one needs to find the  $(\eta, P_F^*)$  relation for all design scenarios under consideration, and the resulting design factors will vary for different design scenarios. In principle, the distribution  $c(\theta)/SR(Z, \theta)$  should depend on  $\theta$  and hence, we state the contrary as a hypothesis in Appendix. The empirical study shows that this hypothesis is reasonable. In this section, we attempt to explain qualitatively why the distribution of  $c(\theta)/SR(Z, \theta)$  does not appear to change drastically with  $\theta$ .

This weak dependency is explained as follows by considering the special case of  $\eta = 0.5$ . In this case,  $c(\theta)$  is similar to the nominal value of  $SR(Z, \theta)$ . Although the distribution of  $SR(Z, \theta)$  may change drastically with  $\theta$  (see Figure 5(a)), the distribution of  $c(\theta)/SR(Z, \theta)$  usually does not (see Figure 5(b)) due to the cancellation effect between  $SR(Z, \theta)$  and the nominal value of  $SR(Z, \theta)$ . The same phenomenon remains for  $\eta \neq 0.5$ . Later in the demonstrating drilled shaft example, the invariance of the  $c(\theta)/SR(Z, \theta)$  distribution over  $\theta$  will be verified empirically.

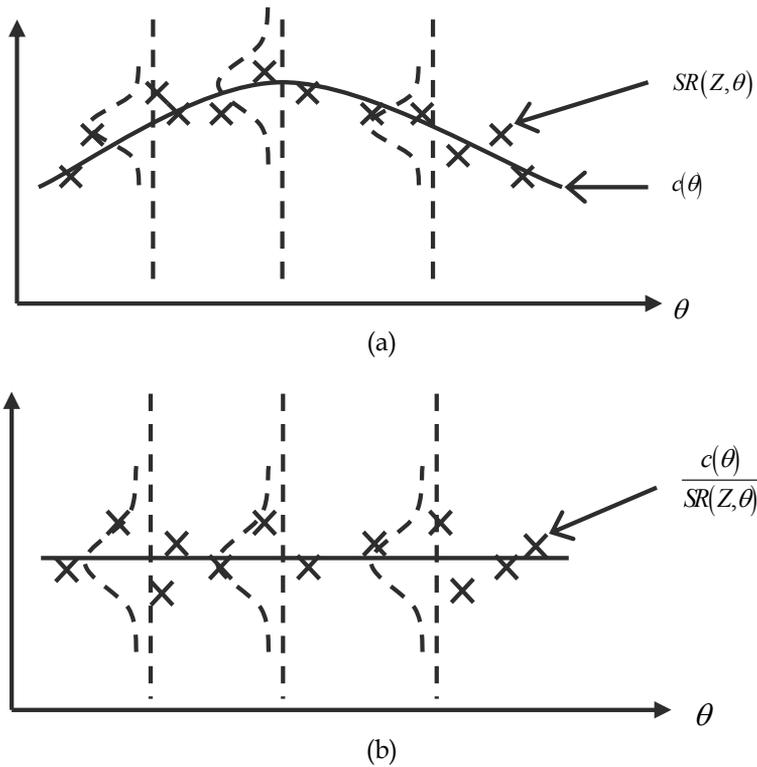


Fig. 5. Illustration of the distributions of  $SR(Z, \theta)$  and  $c(\theta)/SR(Z, \theta)$  (a). Illustration of the distribution of  $SR(Z, \theta)$  (b). Illustration of the distribution of  $c(\theta)/SR(Z, \theta)$

If the distribution of  $c(\theta)/SR(Z, \theta)$  is indeed approximately invariant over  $\theta$ , the relation between  $(\eta, P_F^*)$  can be found by the following equation:

$$P\left(\frac{c(\theta)}{SR(Z, \theta)} > 1\right) = P_F^* \quad (43)$$

where  $\theta$  is treated as random and uniformly distributed over the allowable design region. It is clear that  $P_F^*$  is simply the exceedance probability of  $c(\theta)/SR(Z, \theta)$  over the unity. Therefore, the relation between  $\eta$  and  $P_F^*$  can be determined by any reliability method, in particular the Monte Carlo simulation (MCS): draw  $N$  samples of  $(Z, \theta)$ , where  $Z$  samples are drawn from  $p(z|\theta)$ , and  $\theta$  samples are drawn from the uniform distribution over the allowable design region. Each  $(Z, \theta)$  sample pair can be used to obtain a sample of  $c(\theta)/SR(Z, \theta)$ . At the end of MCS, we have  $N$  samples of  $c(\theta)/SR(Z, \theta)$ . For a chosen  $\eta$  value, the corresponding  $P_F^*$  value can be simply estimated as the ratio that  $c(\theta)/SR(Z, \theta)$  samples are greater 1. By changing the  $\eta$  value and repeating the MCS, one can obtain the entire relation between  $\eta$  and  $P_F^*$

According to Appendix and the above discussions, the reliability constraint Eq. (41) can be transformed into the algebraic design constraint Eq. (40) if the distribution of  $c(\theta)/SR(Z, \theta)$  is invariant over  $\theta$ . Let us denote  $\Sigma_R = \{\theta: P(SR(Z, \theta) < 1 | \theta) \leq P_F^*\}$  be the design region that satisfies the reliability constraint that failure probability is less or equal to the target failure

probability,  $P_F^*$ . From Appendix, if the distribution of  $c(\theta)/SR(Z,\theta)$  is indeed invariant over  $\theta$ , it is assured that the region  $\Sigma_R$  is identical to the following region:  $\Sigma_S = \{\theta: c(\theta) \geq 1\}$ , so the algebraic design formats and RBD become equivalent.

### 4.3 Determining the design factors based on the principle

The following steps can be taken to find the design factors, such as safety factor, load factors, resistance factors, and partial factors, corresponding to any prescribed  $P_F^*$ :

1. Find the relation between  $\eta$  and the design factors:
  - a. For the safety-factor design format, the safety factor  $SF^\eta$  is  $SR(z^*,\theta)/SR^\eta(\theta)$ , where the  $\eta$  quantile  $SR^\eta(\theta)$  can be easily estimated by MCS.
  - b. For LRFD, the resistance factor  $\gamma_S^\eta$  is  $S^\eta(\theta)/S(z^*,\theta)$ , while the load factor  $\gamma_{Lx}^\eta$  is  $L_x^{1-\eta}(\theta)/L_x(z^*,\theta)$ .
  - c. For MRFD, the resistance factor  $\gamma_{Sx}^\eta$  is  $S_x^\eta(\theta)/S_x(z^*,\theta)$ , while the load factor  $\gamma_{Lx}^\eta$  is  $L_x^{1-\eta}(\theta)/L_x(z^*,\theta)$ .
  - d. For the partial-factor design format, the partial factor  $\gamma_i^\eta$  for  $Z_i$  is  $z_i^\eta/z_i^*$  if  $Z_i$  is stabilizing and is  $z_i^{1-\eta}/z_i^*$  if  $Z_i$  is destabilizing.
2. Find the relation between the pair  $(\eta, P_F^*)$  by solving Eq. (43). This has been presented previously, i.e., simulating  $c(\theta)/SR(Z,\theta)$  samples and find the ratio of less than 1. In Eq. (43), the definition of  $c(\theta)$  for various algebraic design formats are different:
  - a. For the safety-factor design format,  $c(\theta) = SR(z^*,\theta)/SF^\eta$ .
  - b. For LRFD,  $c(\theta) = \gamma_S^\eta S(z^*,\theta) / [\gamma_{LD}^\eta L_D(z^*,\theta) + \gamma_{LL}^\eta L_L(z^*,\theta)]$ .
  - c. For MRFD,  $c(\theta) = \Sigma_x[\gamma_{Sx}^\eta S_x(z^*,\theta)] / [\gamma_{LD}^\eta L_D(z^*,\theta) + \gamma_{LL}^\eta L_L(z^*,\theta)]$ .
  - d. For the partial-factor design format,  $c(\theta) = SR(\gamma_1^\eta z_1^*, \gamma_2^\eta z_2^*, \dots, \gamma_p^\eta z_p^*, \theta)$ .
3. Given the prescribed target failure probability,  $P_F^*$ , find the corresponding probability threshold  $\eta$  from the result in Step 1.
4. Once the corresponding probability threshold  $\eta$  is found, the required design factor can be determined accordingly according to the relations presented in Step 1.
  - a. For safety-factor design format, the resulting algebraic design constraint is  $c(\theta) = SR(z^*,\theta)/SF^\eta \geq 1$ .
  - b. For LRFD, the resulting algebraic design constraint is  $c(\theta) = \gamma_S^\eta S(z^*,\theta) / [\gamma_{LD}^\eta L_D(z^*,\theta) + \gamma_{LL}^\eta L_L(z^*,\theta)] \geq 1$ .
  - c. For MRFD, the resulting algebraic design constraint is  $c(\theta) = \Sigma_x[\gamma_{Sx}^\eta S_x(z^*,\theta)] / [\gamma_{LD}^\eta L_D(z^*,\theta) + \gamma_{LL}^\eta L_L(z^*,\theta)] \geq 1$ .
  - d. For partial-factor design format, the resulting algebraic design constraint is  $c(\theta) = SR(\gamma_1^\eta z_1^*, \gamma_2^\eta z_2^*, \dots, \gamma_p^\eta z_p^*, \theta) \geq 1$ .

According to the derivations given in Appendix A, the design based on the algebraic constraint  $c(\theta) \geq 1$  is identical to the probabilistic constraint  $P(SR(Z,\theta) < 1 | \theta) \leq P_F^*$ .

### 4.4 Example

The same drilled shaft problem that was taken in the previous section will be used to demonstrate the RBD. Let the diameter  $B$  and length  $L$  be the two design parameters (i.e.,  $\theta$  contains  $B$  and  $L$ ) that are subjected to change in the design process. Let us further assume there is a practicality design constraints  $0.8 \text{ m} \leq B \leq 1.5 \text{ m}$  and  $49.6 \text{ m} \leq L \leq 74.7 \text{ m}$ . These practicality design constraints are realistic since most drilled shafts have diameters ranging from 0.8 m to 1.5 m and since most drilled shafts may be bottomed in strata with high strengths, in our case, the gravel or rock layer (49.6 m and 74.7 m are the limiting depths of

the gravel and rock layers). Other conditions, such as the ground profile and the mean values and c.o.v.s of  $L_D$  and  $L_L$ , remain the same as in the reliability analysis section. The target failure probability  $P_F^*$  is taken to be 0.001, i.e., the design goal is to adopt a certain combination of B and L so that  $P(SR(Z,\theta) < 1 | \theta) \leq 0.001$ . For brevity, only the detailed steps for MRFD will be demonstrated, but the results for the safety-factor design, LRFD, and partial-factor design will be still presented. Recall that the total resistance S is provided by  $S_c, S_s, S_g$ , and  $S_r$ :

$$\begin{aligned}
 S_c(Z, \theta) &= \pi B e^{2.7+0.3[\ln s_{u,m} - e_{su}] + \varepsilon_{sc}} t_c \\
 S_s(Z, \theta) &= \pi B e^{1.0802-0.6588 \ln(d_s) + [\ln(\sigma'_{vs,m}) - e_{\sigma'_{vs}}] + \varepsilon_{ss}} t_s \\
 S_g(Z, \theta) &= \pi B e^{2.1792-0.7528 \ln(d_g) + [\ln(\sigma'_{vg,m}) - e_{\sigma'_{vg}}] + \varepsilon_{sg}} \min(L - t_c - t_s, t_g) \\
 S_r(Z, \theta) &= \pi B e^{3.0253+0.414[\ln q_{u,m} - e_{qu}] + \varepsilon_{sr}} \min(L - t_c - t_s - t_g, 0)
 \end{aligned} \tag{44}$$

Note that B and L are now subjected to change in the design process;  $\min(A,B)$  is the minimum value among A and B, and the two terms  $\min(L-t_c-t_s, t_g)$  and  $\min(L-t_c-t_s-t_g, 0)$  are there due to the fact that the shaft may not penetrate the entire gravel and rock layers when the length L is not large enough.

Step 1 - determine the relations between  $\eta$  and the MRFD design factors. Note that these relations are independent of the diameter B and the embedment lengths in various strata, i.e.  $t_c, t_s, \min(L-t_c-t_s, t_g)$ , etc. Therefore, these relations can be determined by fixing B and L at any values, for this example, B = 1.2 m and L = 74.7 m. Recall that the resistance factor corresponding to  $\eta$  is  $S_x^\eta(\theta)/S_x^*(\theta)$ , while the load factor corresponding to  $\eta$  is  $L_x^{1-\eta}(\theta)/L_x^*(\theta)$ . The nominal values  $S_x^*$  and those for the dead and live loads are taken to be

$$\begin{aligned}
 S_c^* &= S_c(z^*, \theta) = \pi B e^{2.7+0.3 \ln s_{u,m}} t_c \\
 S_s^* &= S_s(z^*, \theta) = \pi B e^{1.0802-0.6588 \ln(d_s) + \ln(\sigma'_{vs,m})} t_s \\
 S_g^* &= S_g(z^*, \theta) = \pi B e^{2.1792-0.7528 \ln(d_g) + \ln(\sigma'_{vg,m})} \min(L - t_c - t_s, t_g) \\
 S_r^* &= S_r(z^*, \theta) = \pi B e^{3.0253+0.414 \ln q_{u,m}} \min(L - t_c - t_s - t_g, 0) \\
 L_D^* &= \mu_{LD} = 8000 \text{ kN} \quad L_L^* = \mu_{LL} = 4000 \text{ kN}
 \end{aligned} \tag{45}$$

As a result,

$$\begin{aligned}
 \gamma_{S_c}^\eta &= [\eta \text{ quantile of } S_c(Z, \theta)] / S_c(z^*, \theta) = \eta \text{ quantile of } [S_c(Z, \theta) / S_c(z^*, \theta)] \\
 &= \eta \text{ quantile of } e^{\varepsilon_{sc} - 0.3e_{su}} \\
 \gamma_{S_s}^\eta &= \eta \text{ quantile of } e^{\varepsilon_{ss} - e_{\sigma'_{vs}}} \quad \gamma_{S_g}^\eta = \eta \text{ quantile of } e^{\varepsilon_{sg} - e_{\sigma'_{vg}}} \quad \gamma_{S_r}^\eta = \eta \text{ quantile of } e^{\varepsilon_{sr} - 0.414e_{qu}} \\
 \gamma_{L_D}^\eta &= [(1-\eta) \text{ quantile of } L_D(Z, \theta)] / L_D^* = (1-\eta) \text{ quantile of } [L_D(Z, \theta) / 8000] \\
 \gamma_{L_L}^\eta &= (1-\eta) \text{ quantile of } [L_L(Z, \theta) / 4000]
 \end{aligned} \tag{46}$$

Figure 6 shows the relations between  $\eta$  and the MRFD design factors.

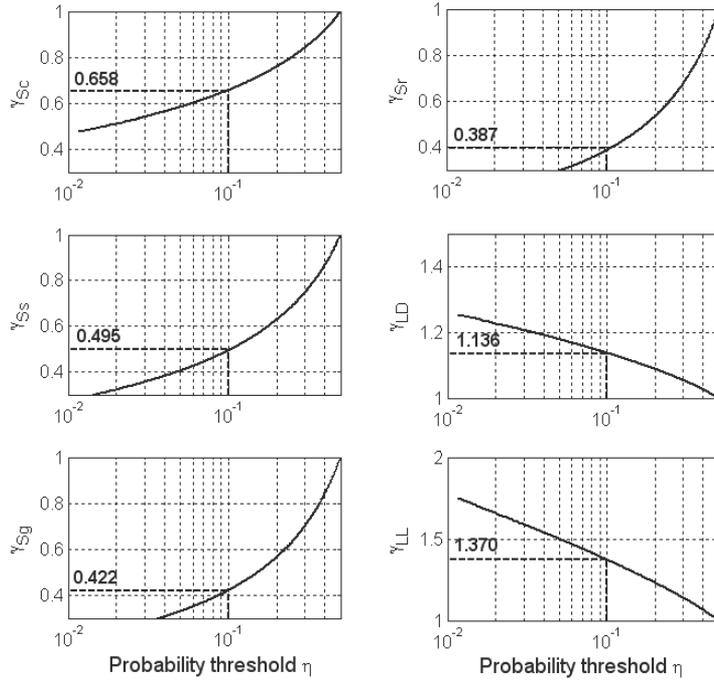


Fig. 6. Relations between  $\eta$  and the MRFD design factors.

Step 2 - find the relation between  $(\eta, P_F^*)$  by solving Eq. (43). For instance, when  $\eta = 0.1$ , the corresponding design factors can be readily from Figure 6 to be  $\gamma_{Sc} = 0.658$ ,  $\gamma_{Ss} = 0.495$ ,  $\gamma_{Sg} = 0.422$ ,  $\gamma_{Sr} = 0.387$ ,  $\gamma_{LD} = 1.136$ , and  $\gamma_{LL} = 1.370$ . Therefore,

$$\frac{c(\theta)}{SR(Z, \theta)} = \frac{[\gamma_{Sc}^\eta S_c^* + \gamma_{Ss}^\eta S_s^* + \gamma_{Sg}^\eta S_g^* + \gamma_{Sr}^\eta S_r^*] \times [L_D(Z, \theta) + L_L(Z, \theta)]}{[S_c(Z, \theta) + S_s(Z, \theta) + S_g(Z, \theta) + S_r(Z, \theta)] \times [\gamma_{LD}^\eta L_D^* + \gamma_{LL}^\eta L_L^*]} \quad (47)$$

$$= \frac{\left[ \begin{aligned} &0.658 e^{2.7+0.3 \ln s_{u,m}} t_c + 0.495 e^{1.0802-0.6588 \ln(d_s) + \ln(\sigma'_{vs,m})} t_s \\ &+ 0.422 e^{2.1792-0.7528 \ln(d_g) + \ln(\sigma'_{vg,m})} \min(L - t_c - t_s, t_g) \\ &+ 0.387 e^{3.0253+0.414 \ln q_{u,m}} \min(L - t_c - t_s - t_g, 0) \end{aligned} \right] \cdot [L_D(Z, \theta) + L_L(Z, \theta)]}{\left[ \begin{aligned} &e^{2.7+0.3 [\ln s_{u,m} - e_{su}]} + \epsilon_{Sc} t_c + e^{1.0802-0.6588 \ln(d_s) + [\ln(\sigma'_{vs,m}) - e_{\sigma'_{vs}}]} + \epsilon_{Ss} t_s \\ &+ e^{2.1792-0.7528 \ln(d_g) + [\ln(\sigma'_{vg,m}) - e_{\sigma'_{vg}}]} + \epsilon_{Sg} \min(L - t_c - t_s, t_g) \\ &+ e^{3.0253+0.414 [\ln q_{u,m} - e_{qu}]} + \epsilon_{Sr} \min(L - t_c - t_s - t_g, 0) \end{aligned} \right] \cdot [1.136 \times 8000 + 1.370 \times 4000]}$$

According to Eq. (43), the corresponding  $P_F^*$  is exactly  $P(c(\theta)/SR(Z, \theta) > 1)$ . Monte Carlo simulation can be taken to simulate many  $c(\theta)/SR(Z, \theta)$  samples estimate, and the estimate of

the probability of exceeding 1 is exactly the  $P_F^*$  value corresponding to  $\eta = 0.1$ . Note that for the MCS, the design parameters  $B$  and  $L$  should be taken to be random and uniformly distributed over the allowable design region  $0.8 \text{ m} \leq B \leq 1.5 \text{ m}$  and  $49.6 \text{ m} \leq L \leq 74.7 \text{ m}$ . The entire  $(\eta, P_F^*)$  relation can be obtained by changing the probability threshold  $\eta$  and conduct the same MCS. Figure 7 shows the resulting relation between  $(\eta, P_F^*)$ . Since the target failure probability  $P_F^*$  is 0.001, the required probability threshold should be 0.071 (see Figure 7). By inverting Figure 6 with  $\eta = 0.071$ , it can be found that  $\gamma_{sc} = 0.618$ ,  $\gamma_{ss} = 0.446$ ,  $\gamma_{sg} = 0.372$ ,  $\gamma_{sr} = 0.336$ ,  $\gamma_{LD} = 1.158$ , and  $\gamma_{LL} = 1.436$ . These are the MRFD factors that should be taken for a RBD with  $P_F^* = 0.001$ , and the resulting algebraic design constraint is

$$c(\theta) = \frac{0.618 \times S_c^* + 0.446 \times S_s^* + 0.372 \times S_g^* + 0.336 \times S_r^*}{1.158 \times 8000 + 1.436 \times 4000} \geq 1 \quad (48)$$

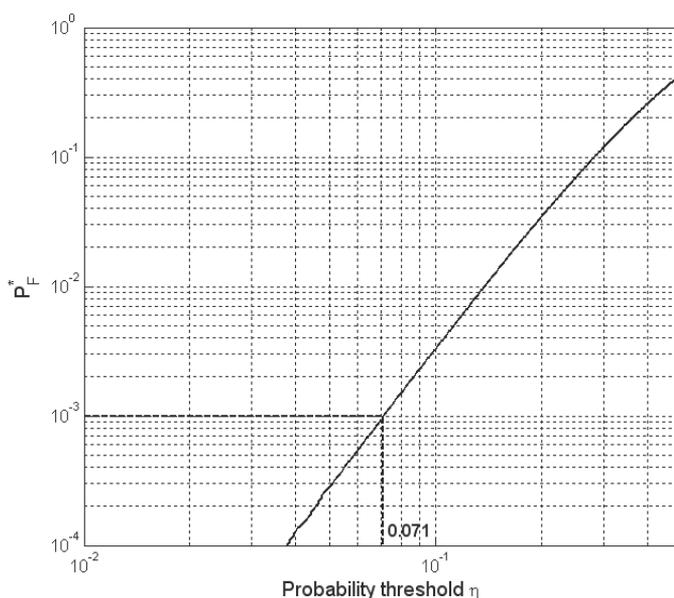


Fig. 7. Relation between  $(\eta, P_F^*)$

To examine the robustness of the resulting MRFD design factors for  $P_F^* = 0.001$ , the following approach is taken. In the allowable design region  $0.8 \text{ m} \leq B \leq 1.5 \text{ m}$  and  $49.6 \text{ m} \leq L \leq 74.7 \text{ m}$ , each of the coordinate axes is discretized into discrete points, creating grid points. MCS with a very large sample size is then conducted at each grid point, giving each point an independent estimate of the failure probability  $P(\text{SR}(Z, \theta) < 1 | \theta)$ . The dividing boundary for  $P(\text{SR}(Z, \theta) < 1 | \theta)$  less and greater than 0.001 is plotted shown in Figure 8. Therefore, the region above the boundary is exactly the allowable reliability design set  $\Sigma_R = \{\theta: P(\text{SR}(Z, \theta) < 1 | \theta) \leq P_F^*\}$ . On the other hand, the allowable design region for the algebraic constraint, i.e.,  $\Sigma_S = \{\theta: c(\theta) \geq 1\}$ , can be also found and is marked with label 'o'. For the unsatisfactory region with  $c(\theta) < 1$ , it is marked with label 'x'. Such comparisons are made in Figure 8, not only for the MRFD design format but also for the other three aforementioned design formats, although the detailed steps of those three design formats are not presented.

It can be seen from Figure 8 that the two sets match one another reasonably well for all four design methods.

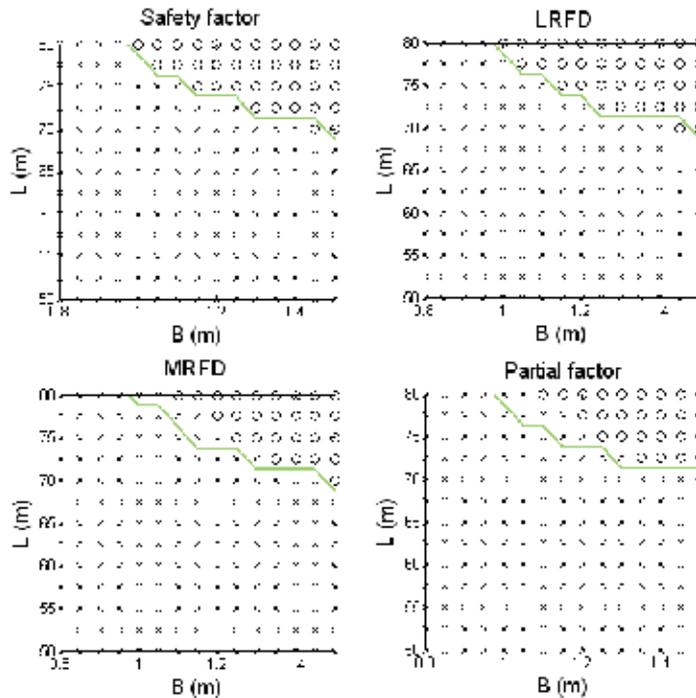


Fig. 8. Comparison between algebraic design constraints and rigorous RBD

#### 4.5 Iterative design process

Given the required MRFD factors  $\gamma_{Sc} = 0.618$ ,  $\gamma_{Ss} = 0.446$ ,  $\gamma_{Sg} = 0.372$ ,  $\gamma_{Sr} = 0.336$ ,  $\gamma_{LD} = 1.158$ , and  $\gamma_{LL} = 1.436$ , the goal now is to design the diameter  $B$  and length  $L$  of the drilled shaft so that  $c(\theta) \geq 1$ . The resulting design should also satisfy failure probability less than 0.001. The design process is iterative. Let us start with a design with diameter = 1.2 m and depth = 70 m. Based on this design dimension and also the information in Table 1, the nominal resistance can be computed from Eq. (45):  $S_c^* = 8448$  kN,  $S_s^* = 2345$  kN,  $S_g^* = 14808$  kN and  $S_r^* = 389$  kN. Note that when calculating  $S_c^*$ ,  $S_s^*$ ,  $S_g^*$ ,  $S_r^*$ , the characteristic values of the geotechnical parameters must be fixed at the measured values (or average values) of those parameters. Now compute

$$c(\theta) = \frac{\gamma_{Sc}^{\eta} S_c^* + \gamma_{Ss}^{\eta} S_s^* + \gamma_{Sg}^{\eta} S_g^* + \gamma_{Sr}^{\eta} S_r^*}{\gamma_{LD}^{\eta} \cdot L_D^* + \gamma_{LL}^{\eta} \cdot L_L^*} \quad (49)$$

$$= \frac{0.618 \cdot 8448 + 0.446 \cdot 2345 + 0.372 \cdot 14808 + 0.336 \cdot 389}{1.158 \cdot 8000 + 1.436 \cdot 4000} = 0.793 < 1$$

It is therefore concluded that the design is not satisfactory for target failure probability less than 0.001. A design with greater length or larger diameter is needed, and iterations should be taken until  $c(\theta)$  is greater than 1.

## 5. Conclusion

Monte Carlo based reliability analysis methods and reliability-based design methods are introduced in this chapter. A realistic geotechnical design example is developed and is used to demonstrate the uses of all methods. The main benefits for the Monte Carlo based methods include (a) their implementations do not require the knowledge of optimization skills, as required by the first-order reliability methods; and (b) they are mostly general and robust to the dimension of random variables and problem complexity, hence ideal for geotechnical problems. One possible drawback is that these Monte Carlo based methods are more time consuming, but this issue has been alleviated greatly due to recent development of powerful personal computers. As a result, these methods are believed to be ideal for practical implementations.

## 6. Appendix derivations for the equivalence principle

The safety ratio  $SR(Z,\theta)$  is a positive-valued random variable, taking values strictly larger than 0, i.e.,  $SR(Z,\theta) > 0$ , and  $c(\theta)$  is a positive-valued deterministic function of  $\theta$ .

$$\begin{aligned}
 &P(SR(Z,\theta) < 1 | \theta) \\
 &= P(0 < SR(Z,\theta) < 1 | \theta) \\
 &= P(1/SR(Z,\theta) > 1 | \theta) && \text{because } SR > 0 \\
 &= P(c(\theta)/SR(Z,\theta) > c(\theta) | \theta) && \text{because } c > 0
 \end{aligned} \tag{50}$$

Observe that  $P(c(\theta)/SR(Z,\theta) > x | \theta)$  is a non-increasing (i.e. equal or decreasing) function with  $x$  (a constant), since  $P(c(\theta)/SR(Z,\theta) > x | \theta) = P(SR(Z,\theta) < c(\theta)/x | \theta)$  is cumulative distribution function of  $SR(Z,\theta)$  which is a non-decreasing (i.e. equal or increasing) function with  $c(\theta)/x$  by definition.

Hence, if  $c(\theta) \geq 1$ ,

$$P(c(\theta)/SR(Z,\theta) > c(\theta) | \theta) \leq P(c(\theta)/SR(Z,\theta) > 1 | \theta) \tag{51}$$

If we further let

$$P(c(\theta)/SR(Z,\theta) > 1 | \theta) = P_F^* \tag{52}$$

Then,

$$P(SR(Z,\theta) < 1 | \theta) = P(c(\theta)/SR(Z,\theta) > c(\theta) | \theta) \leq P(c(\theta)/SR(Z,\theta) > 1 | \theta) = P_F^* \tag{53}$$

Note that the hypothesis is required for practicality. The above constitutes a proper proof of equivalence if we allow  $\eta$  to be a function of  $\theta$ . In summary, the practical usefulness of the equivalence between the statement  $\{\theta: c(\theta) \geq 1\}$  and  $\{\theta: P(SR(Z,\theta) < 1 | \theta) \leq P_F^*\}$  is predicated on the possibility of finding a  $\eta$  which is *not* a function of  $\theta$  for any prescribed target failure probability,  $P_F^*$ .

## 7. References

- Ang, A.H.-S. & Tang, W.H. (1984). *Probability Concepts in Engineering Planning and Design, Vol. I: Basic Principles*. John Wiley & Sons.
- Au, S. K., Papadimitriou, C. & BECK, J.L. (1999). Reliability of uncertain dynamical systems with multiple design points. *Structural Safety*, 21, 113-133.
- Au, S.K. & Beck, J.L. (2001). Estimation of small failure probability in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, 16, 263-277.
- Au, S.K. & Beck, J.L. (2003). Importance sampling in high dimensions. *Structural Safety*, 25(2), 139-163.
- Ching, J. & Phoon, K.K. (2010), Quantile framework for simplified geotechnical reliability-based design. To appear in *Proceedings of 2011 Georisk Conference*.
- Ching, J., Beck, J.L. & Au, S.K. (2005). Hybrid subset simulation method for reliability estimation of dynamic systems subject to stochastic excitations. *Probabilistic Engineering Mechanics*, 20(3), 199-214.
- Der Kiureghian, A. (2000). The geometry of random vibrations and solutions by FORM and SORM. *Probabilistic Engineering Mechanics*, 15, 81-90.
- Der Kiureghian, A. & Dakessian, T. (1998). Multiple design points in first and second-order reliability. *Structural Safety*, 20, 37-49.
- Ghanem, R. & Spanos, P. (1991). A spectral stochastic finite element formulation for reliability analysis. *Journal of Engineering Mechanics, ASCE*, 117(10), 2351-2372.
- Hasofer, A.M. & Lind, N.C. (1974). Exact and invariant second-moment code format. *Journal of Engineering Mechanics, ASCE*, 100(1), 111-121.
- Homenbichler, M. & Rackwitz, R. (1988). Improvement of second-order reliability estimates by importance sampling. *Journal of Engineering Mechanics, ASCE*, 114(12), 2195-2198.
- Liu, P.L. & Der Kiureghian, A. (1991). Optimization algorithms for structural reliability. *Structural Safety*, 9(3), 161-177.
- Melchers, R.E. (1989). Importance sampling in structural systems. *Structural Safety*, 6, pp.3-10.

# A Monte Carlo Framework to Simulate Multicomponent Droplet Growth by Stochastic Coalescence

Lester Alfonso<sup>1</sup>, Graciela Raga<sup>2</sup> and Darrel Baumgardner<sup>2</sup>

<sup>1</sup>*Universidad Autónoma de la Ciudad de México, Mexico City,*

<sup>2</sup>*Universidad Nacional Autónoma de México  
México*

## 1. Introduction

The accurate modeling of the interactions between aerosols and cloud droplets for a multi-component system is a very difficult task in cloud modeling, since to express a variety of properties of the hydrometeors (such as the masses of water and soluble materials inside droplets) there is a need for multi-dimensional size distributions.

The aerosol distribution becomes important as the cloud drops evaporate and the solutes are recycled into aerosols that can serve as cloud condensation nuclei (CCN): the larger the mass of a hygroscopic aerosol, the lower the supersaturation needed to form a cloud droplet. In the marine environment, the aerosol recycling process is believed to be the major mechanism responsible for the bimodal shape of the aerosol size distributions (Flossmann, 1994; Feingold and Kreidenweiss, 1996). The heterogeneous chemical reactions, which add nonvolatile solute to each cloud droplet, strongly depend on the salt content and pH of the droplet (Alfonso and Raga, 2004). Since aerosols also have a significant influence on cloud microphysics and cloud radiative properties, it is necessary to simulate aerosol processes realistically and with adequate accuracy.

The usual approach adopted in detailed cloud microphysical modeling is to describe the aerosols and drops in two separate one-dimensional size distributions. Within this approach, only the average aerosol mass contained in drops of certain size is known, and it is not possible to accurately track the aerosol mass distribution within cloud droplets (Jacobson, 1999).

For the deterministic case (based on the solution of the kinetic collection or stochastic collection equation), the aerosol processing due to collision-coalescence was addressed by Bott (2000) by extending his previous model (Bott, 1998) to two-dimensional distributions. Within this framework each particle is characterized both by the mass of its dry aerosol nucleus and by its water mass. By adopting this framework, there is no need to parameterize the activation process.

Nevertheless, in real situations, there are several types of aerosols that act as CCN, and form an internal or an external mixture. Thus, the number of components of the system can be larger than two. The solution of the kinetic collection equation when the number of

components is larger than two is not an easy task and the alternative seems to be the stochastic treatment of the coalescence process (Alfonso et al., 2009).

Clouds that also contain ice crystals, as well as aerosols and cloud droplets, constitute even much more complex systems to be modeled. For mixed phase clouds, the components of the system are not only the particle mass and the aerosol mass inside the particles, but also the type of ice particle, such as ice crystals of different geometries (columns, plates, and dendrites), graupel or aggregates. In this case, the number of components in the system is very large and the kinetic framework is extremely difficult to implement. As a consequence, simplified treatments are adopted to deal with this problem. For example, in many models only one type of ice crystal is considered in order to make the problem more manageable.

Therefore, most models do not deal with several types of ice, and not take into account the aggregation of ice particles. When a variety of types of geometries are considered, then the system of kinetic equations can be actually very complex. For example, Khain and Sednev (1995) considered the interactions between water drops, columns, crystals (plate like crystals and dendrites), snowflakes, graupel and hail. The resulting system consists of seven complex kinetic equations that need to be solved with the Berry and Reinhardt (1974) method. In Alfonso et. al (2009), the algorithm of Gillespie (1976) for chemical reactions in the formulation proposed by Laurenzi et al. (2002) was applied to calculate the evolution of a two-component system (the masses of pure water and soluble material). The algorithm could be easily extended to any multi-component cloud system, with the possible inclusion of the ice phase.

Another less known drawback of the deterministic approach (based on the solution of the kinetic collection equation) is the fact that this equation can exhibit non-conservation of mass (gelation) under certain conditions. These limitations of the KCE are carefully analyzed in two previous papers (Alfonso et al, 2008 and Alfonso et al., 2010) by a direct comparison of numerical and analytical solutions of the KCE with true averages obtained with the stochastic method of Gillespie (1976). In these papers, a numerical criterion is proposed in order to calculate the validity time or breakdown time of the KCE.

Although it is easy to implement, the stochastic framework developed by Gillespie (1976) has an important limitation: It is computationally very expensive, and consequently, only small cloud volumes can be considered in the simulations. A possible solution to this problem can rely in the implementation of the grouping method (Ormel and Spaan, 2008) that allows modeling coalescence in a sufficiently large region.

This chapter is organized as follows. In section 2 we are concerned with drawbacks of the deterministic framework. Section 3 describes the multi-component collection stochastic algorithm and its application to solve kinetic collection equation. In Section 4 the multi-component stochastic algorithm is incorporated into a particle based microphysical model, and applied to model the microphysical evolution of an orographic cloud in Section 5. Section 6 summarizes the main results of the chapter.

## **2. Drawbacks of the deterministic approach**

### **2.1 Non conservation of mass after gelation**

One of the most important mechanisms for the formation of rain is the collision and coalescence of smaller droplets into larger ones. The deterministic approach to model this process is based in the solution of the kinetic collection (stochastic collection, coagulation) equation, which in discrete form is expressed as (Pruppacher and Klett, 1997):

$$\frac{\partial N(i,t)}{\partial t} = \frac{1}{2} \sum_{j=1}^{i-1} K(i-j,j)N(i-j)N(j) - N(i) \sum_{j=1}^{\infty} K(i,j)N(j) \tag{1}$$

where  $N(i,t)$  is the average number of droplets with mass  $x_i$  as a function of time. In Eq. (1), the time rate of change of the average number of droplets with mass  $x_i$  is determined as the difference between two terms: the first term describes the average rate of production of droplets of mass  $x_i$  due to coalescence between pairs of drops whose masses add up to mass  $x_i$ , and the second term describes the average rate of depletion of droplets with mass  $x_i$  due to their collisions and coalescence with other droplets.

The known limitations of the KCE are analyzed carefully in two papers (Alfonso et al., 2008 and Alfonso et al., 2010) by a direct comparison of numerical and analytical solutions of the KCE with true averages obtained with the stochastic method of Gillespie (1976). In these papers, a numerical criterion is proposed in order to calculate the validity time or breakdown time of the KCE.

The collision coalescence process is a stochastic one and is more accurately described by the stochastic coagulation equation for the joint probability distribution  $P(n_1, n_2, \dots, n_k, \dots, t)$  for the occupation numbers  $\bar{n}=(n_1, n_2, \dots, n_k, \dots)$  at time  $t$ . This equation has the form (Bayewitz et al., 1974; Lushnikov, 1978; Tanaka and Nakazawa, 1993; Inaba et al., 1999; Wang et al., 2006):

$$\begin{aligned} \frac{\partial P(\bar{n})}{\partial t} = & \sum_{i=1}^N \sum_{j=i+1}^N K(i,j)(n_i+1)(n_j+1)P(\dots, n_i+1, \dots, n_j+1, \dots, n_{i+j}-1, \dots; t) \\ & + \sum_{i=1}^N \frac{1}{2} K(i,i)(n_i+2)(n_i+1)P(\dots, n_i+2, \dots, n_{2i}-1, \dots; t) \\ & - \sum_{i=1}^N \sum_{j=i+1}^N K(i,j)n_i n_j P(\bar{n}; t) - \sum_{i=1}^N \frac{1}{2} K(i,i)n_i(n_i-1)P(\bar{n}; t) \end{aligned} \tag{2}$$

In (2)  $n_i$  is the number of droplets with mass  $x_i$ , and  $N$  is the total number of size bins. The KCE results from taking the first moments::

$$\langle n_k \rangle = \sum_{\bar{n}} n_k P(\bar{n}; t) \tag{3}$$

and assuming that  $\langle n_i n_j \rangle = \langle n_i \rangle \langle n_j \rangle$ . Under these assumptions Eq.(2) reduces to the kinetic collection equation (1). Then, the average spectrum obtained from Eq.(1), and the ensemble average obtained from different realizations of the stochastic collection process are different. Bayewitz et al. (1974) showed that the solution of the KCE and the expected values calculated from the stochastic equation are equal only if the covariances are omitted from the probabilistic model.

Equation (1) is not expected to be accurate when the initial number of particles is small, or if  $K(i,j)$  increases sufficiently rapidly with  $x_i$  and  $x_j$ . For example, in the analytic solution for the case  $K(i,j)=Cx_i \times x_j$ , with monodisperse initial conditions  $N(1,0)=N_0$ , the total mass starts to decrease after a certain time and there is an increase in the second moment. Drake (1972) calculated the analytical solutions of the KCE for polynomials of the form  $K(x,y)=Cx y$ . In this case the second moment evolution is given by:

$$M_2(\tau) = \frac{M_2(t_0)}{1 - CM_2(t_0)\tau} \quad (4)$$

Where  $M_2(\tau)$  defined by the expression (for the discrete case):

$$M_2(t) = \sum_{i=1}^{N_d} x_i^2 N(i,t) \quad (5)$$

Where  $N_d$  is the total number of size bins. Note that when

$$\tau = [CM_2(t_0)]^{-1} \quad (6)$$

The second moment  $M_2$  is undefined and the total mass of the system starts to decrease. This is usually interpreted to mean that a macroscopic runaway particle has formed (known as a gel). The product kernel  $K(i,j) = Cx_i \times x_j$ , is the prototype example where the process exhibits a phase transition (also called gelation). After gelation occurs, there is a transition from a continuous system to one with a continuous distribution *plus* a massive runaway particle. The critical time is defined in terms of the existence of solutions of the coagulation equation (1) which is mass-conserving.  $T_{gel}$  is the largest time such that the discrete model has a solution with  $M_1(t) = M_1(0)$  for  $t < T_{gel}$  and  $M_1(t) < M_1(0)$  for  $t > T_{gel}$ .

As analytical expressions for the gelation time only exist for very simple kernels, it can be estimated (for real kernels relevant to cloud physics) using a Monte Carlo method (Alfonso et al., 2008, 2010), based on the original algorithm proposed by Gillespie (1976). Following the conjecture made by Inaba et al. (1999), the  $T_{gel}$  is estimated as the maximum of the ratio of the standard deviation for the largest particle mass over all the realizations, to the averaged value evaluated from the realizations of the stochastic algorithm:  $M_{L1,S} = STD(M_{L1}) / M_{L1}$ . The standard deviation for the largest droplet mass is calculated for each time by using the expression:

$$STD(M_{L1}) = \sqrt{\frac{1}{N_r} \left( \sum_{i=1}^{N_r} (M_{L1}^i - M_{L1})^2 \right)} \quad (7)$$

where  $M_{L1}$  is the ensemble mean of the mass of the largest droplet over all the realizations (given by Eq. (17)),  $N_r$  is the number of realizations of the Monte Carlo algorithm and  $M_{L1}^i$  is the largest droplet for each realization.

In Alfonso et al. (2008) the gel transition time  $\tau$  for an initial monodisperse distribution of 100 droplets of 14  $\mu\text{m}$  in radius (droplet mass  $1.1494 \times 10^{-8}\text{g}$ ) was estimated from analytical solutions and from Monte Carlo simulations. The volume of the cloud was set equal to 1  $\text{cm}^3$ . Using a value of  $C = 5.49 \times 10^{10} \text{cm}^3 \text{g}^{-2} \text{s}^{-1}$ , then  $\tau$  in (6) is 1379 sec. For the same analytical conditions, the behavior of  $M_{L1,S}$  was calculated from 1000 realizations ( $N_r = 1000$ ) of the Monte Carlo algorithm. The results are displayed in Fig. 1. The maximum of  $\sigma_L$  was obtained for  $\tau = 1335$  sec, very similar to the analytical estimation from Eq. (6), indicating that the statistic parameter in Eq. (7) is a good estimate of the transition to gel.

For more realistic kernels, relevant to cloud physics modeling, the validity time can be estimated in a similar manner. Alfonso et al. (2010) estimated the breakdown of the coagulation equation for the hydrodynamic kernel:

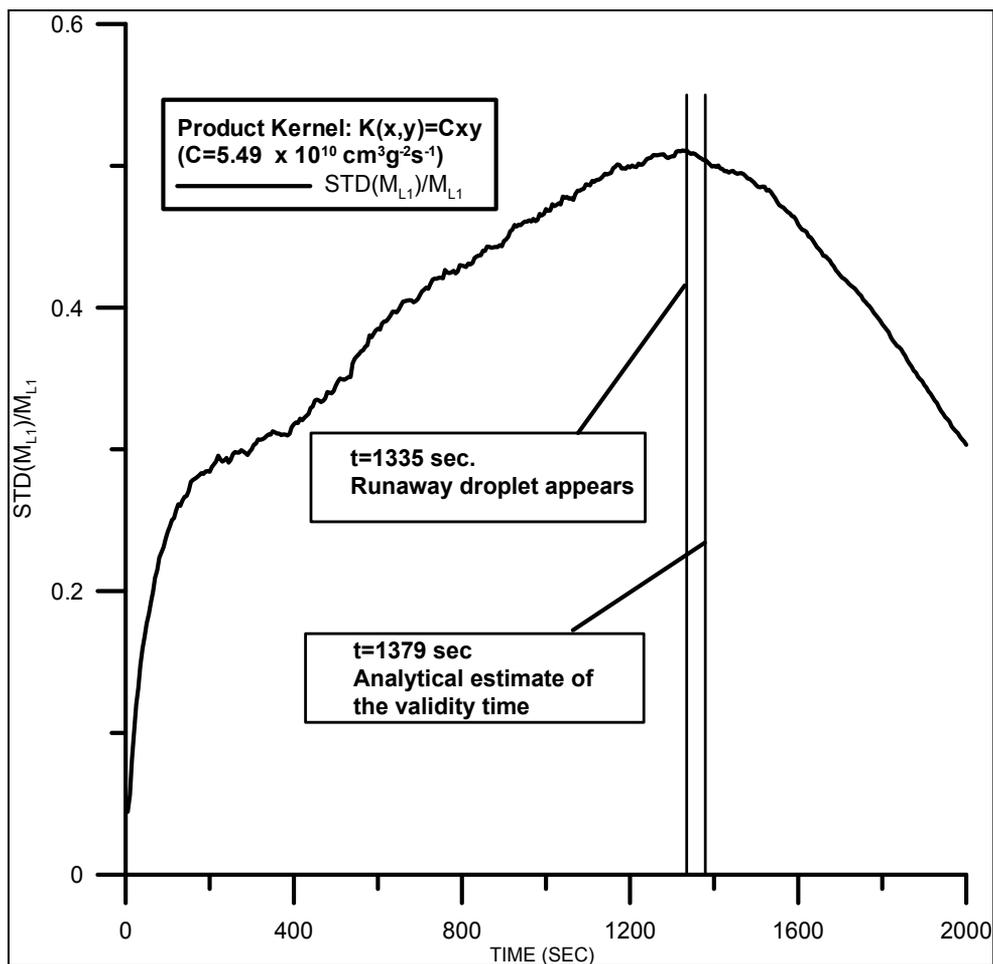


Fig. 1. The ratio (defined in Eq. 7) as a function of time, for the product kernel  $K(x,y)=Cxy$ , ( $C=5.49 \times 10^{10} \text{ cm}^3 \text{ g}^{-2} \text{ s}^{-1}$ ). Note that  $\text{STD}(M_{L1})/M_{L1}$  reaches a maximum when the runaway droplet appears.

$$K(x,y)=\pi[R(x)+r(y)]^2 E(x,y)[V(x)-V(y)], \quad x \geq y \quad (8)$$

where  $V(x), V(y)$  and  $R(x), r(y)$  are the terminal velocities and radiuses of droplets with masses  $x$  and  $y$  respectively, and the values of the collision efficiencies  $E(x,y)$  were taken from Hall (1980). The behavior of the ratio  $M_{L1,S}$  (Eq. 7) was evaluated from 1000 realizations of the Monte Carlo algorithm, and the time when the maximum of the statistics (7) was reached compared with the time when the liquid water content (LWC), obtained numerically with a finite difference scheme, starts to decrease.

A cloud volume of  $1 \text{ cm}^3$  was simulated, that initially contained a bidisperse droplet distribution: 50 droplets of  $14 \mu\text{m}$  in radius, and 50 droplets of  $17.6 \mu\text{m}$  in radius. Figure 2 shows that the liquid water content (or total mass) of the system is no longer conserved after 800 sec. This time is very close to the time when the statistics  $M_{L1,S}$  determined from the Monte Carlo realizations, reaches its maximum (850 sec).

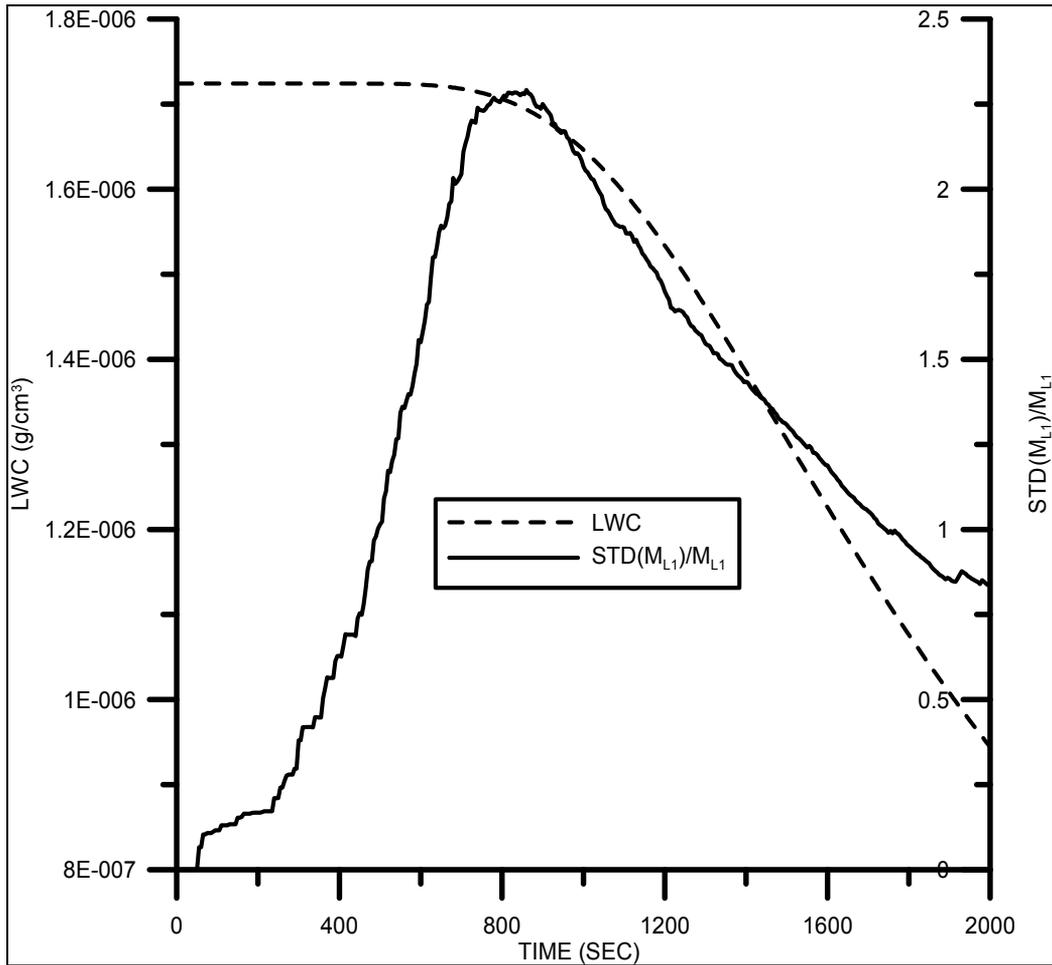


Fig. 2. Time evolution of total liquid water content calculated from the numerical solution of the KCE for the hydrodynamic kernel (dashed line) and the statistics  $STD(M_{L1})/M_{L1}$  (solid line) estimated from the Monte Carlo algorithm. The simulations were performed for the hydrodynamic kernel with a bidisperse initial condition  $N(1;0)=50$  and  $N(2;0)=50$ .

A second simulation was performed, with twice the initial number of droplets, and again the results show a good correspondence between the time of the  $M_{L1,S}$  maximum (430 sec.) and the gelation time obtained from the numerical solution of the KCE (415 sec.). These results confirm the fact that total mass calculated assuming a continuous droplet distribution starts to decrease around the time when the runaway droplet appears.

### 3. Stochastic approach for the collection process

#### 3.1 Definition of species and multi-component stochastic collection algorithm

Within the stochastic framework, each species represents a large number of hydrometeors with the same attributes and position. These attributes are: a) the type of particle

(unactivated and activated droplets, and ice crystals of different geometries), b) the particle mass, and c) the dry aerosol mass for each substance.

Warm clouds are composed of only one type of hydrometeor since unactivated CCN in equilibrium at a given supersaturation and activated droplets are treated as particles of the same type. In this case the attributes are only the droplet mass and the mass of dry aerosols. In a multi-component system, each species is characterized by a vector of properties  $\bar{u}_\mu = (u_1, u_2, \dots, u_N)$ , such that, a droplet with composition  $\bar{u}_\mu$  is a member of the  $\mu$ th species. After time  $t=0$  the species will randomly coalesce according to:

$$A_{u_1, u_2, \dots, u_N} + B_{u'_1, u'_2, \dots, u'_N} = C_{u_1 + u'_1, u_2 + u'_2, \dots, u_N + u'_N} \tag{9}$$

where  $A_{u_1, u_2, \dots, u_N}$  and  $B_{m'_d, m'_{a1}, m'_{a2}, m'_{a3}}$  are droplets with compositions  $\bar{u}_\mu = (u_1, u_2, \dots, u_N)$  and  $\bar{u}'_\mu = (u'_1, u'_2, \dots, u'_N)$ , respectively. The transition probabilities for coalescence events follow Laurenzi et al. (2002) and are given by:

$$a(i,j) = V^{-1} K(i,j) N_i N_j dt \equiv \text{Pr}\{ \text{Probability that two particles of species } i \text{ and } j \text{ (for } i \neq j \text{) with populations (number of particles) } N_i \text{ and } N_j \text{ will collide within the imminent time interval} \} \tag{10}$$

$$a(i,i) = V^{-1} K(i,i) \frac{N_i(N_i-1)}{2} dt \equiv \text{Pr}\{ \text{Probability that two particles of the same species } i \text{ with population (number of particles) } N_i \text{ collide within the imminent time interval} \} \tag{11}$$

In (10) and (11),  $K(i,j)$  is the collection kernel,  $V$  is the cloud volume; and  $N_i$  and  $N_j$  are the total number of droplets for the species  $i$  and  $j$ . An index is assigned to each species (particles with a specific  $\bar{u}_\mu = (u_1, u_2, \dots, u_N)$  composition). Within this framework, there is a unique index  $\nu$  for each pair of droplets  $i, j$  that may collide. For a system with  $N_s$  species  $(S_1, S_2, \dots, S_N)$   $\nu \in \frac{N_s(N_s+1)}{2}$ . The set  $\{\nu\}$  defines the total collision space, and is equal to the total number of possible interactions. The transition probabilities (10) and (11) are then represented by one index ( $a_\nu$ ).

In Alfonso et al. (2009) the stochastic algorithm of Laurenzi et al. (2002) was implemented to calculate two-component droplet growth. This version of the algorithm is difficult to implement in cloud microphysical models, that considered a constant time step. Consequently, a modification is introduced following Sue et al. (2007). First the number of collisions occurring during a time step  $\Delta t$  is determined from the expression:

$$C_T = \frac{\Delta t \sum_{i=1}^{N_s} \sum_{j=1}^{N_s} K(i,j) N_i N_j}{V} \tag{12}$$

Where  $N_s$  is the total number of species. Then, the collision pairs are selected by generating  $C_T$  random numbers  $r_i$  from a uniform distribution in the interval  $(0, 1)$ , and the indexes  $\nu$  for the  $C_T$  collisions determined from the inequality:

$$\sum_{v=1}^{\mu-1} a_v < r_i \alpha < \sum_{v=1}^{\mu} a_v \quad (13)$$

Where  $\alpha = \frac{N_s(N_s+1)}{\sum_{v=1}^2 a_v}$  (14) and  $i=1, \dots, C_T$ . After each collision event, the size distribution is updated by taking into account:

$$N_i = N_i - 1, \quad N_j = N_j - 1 \quad (14)$$

If new species are created, then ( $N_s = N_s + 1$ ). For the new species, the droplet and aerosol masses for each component are equal to the sum of the droplet and aerosol masses of the colliding droplets following Eq. (9). The Monte Carlo algorithm can be summarized as follows:

1. At  $t=0$ , the event counter is set to zero and the initial number of species  $N_1, N_2, \dots, N_N$  is defined.
2. The total number ( $C_T$ ) of collisions in the time interval  $\Delta t$  are determined from the expression (12).
3. Generate  $C_T$  random numbers from a uniform distribution and determine the  $N_T$  collision indexes from (13).
4. Change the numbers of species to reflect the execution of  $C_T$  collisions.
5. Return to step 2.

The approach follow by Sun et al. (2007) was adopted and only a single stochastic experiment was run. This can be justified by considering that the statistical error is proportional to  $1/\sqrt{N_{\text{droplets}}}$ , where  $N_{\text{droplets}}$  is the total number of droplets in the coalescence volume. Then, in order to reduce the statistical error, volumes larger than  $10^3 \text{ cm}^3$  are considered in our simulations. This problem was carefully studied by Laurenzi et al. (2002). They found that the differences between the KCE and the results of the Monte Carlo for one single realization were almost negligible for sufficiently large coalescence volumes.

In order to check the performance of the previously described Monte Carlo algorithm, a simulation with the sum kernel ( $K(i,j) = B(x_i + x_j)$ ) was performed and compared with the analytical solution of the one-component kinetic collection equation derived by Scott (1968) for a monodisperse initial condition.

$$N(t) = N_0(1 - T) \quad (15a)$$

$$T = 1 - \exp(-BN_0 v_0 t) \quad (15b)$$

In Eqs. 15a,b  $N_0$  and  $v_0$  correspond to the initial number and volume of droplets, respectively. We simulate a cloud volume equal to  $5000 \text{ cm}^3$ , containing  $5 \times 10^5$  droplets ( $N_0$ ) of  $14 \mu\text{m}$  in radius ( $v_0 = 1.1494 \times 10^{-8} \text{ cm}^3$ ). Following Long (1974), a value of  $8.83 \times 10^2 \text{ cm}^3 \text{ g}^{-1} \text{ s}^{-1}$  was assumed for constant  $B$  in the sum kernel. The results obtained for the total concentration can be checked in Fig. 3, with a good correspondence between the analytical and the Monte Carlo results.

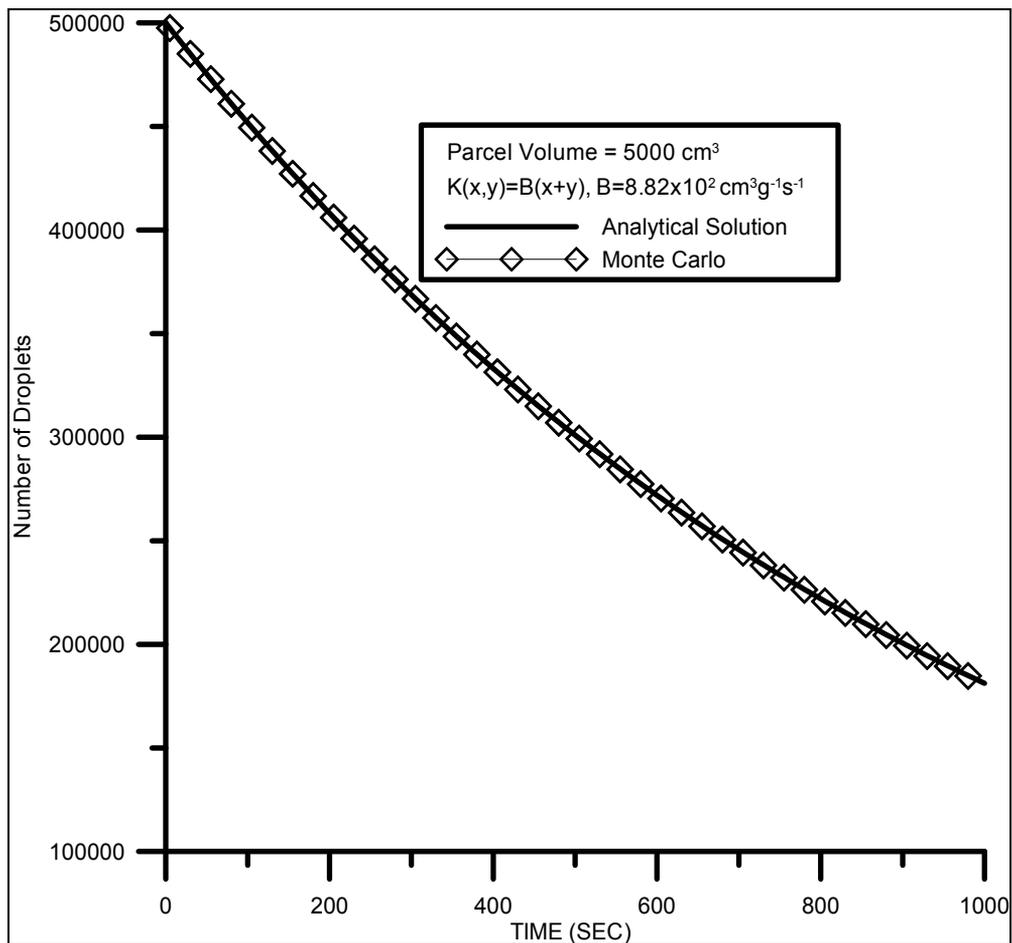


Fig. 3. Time evolution of the total number of particles obtained from the MC method (diamonds), versus the analytical solution of the kinetic collection equation (solid line).

### 3.2 The grouping method

Since the total collision rate  $C_T$  (see equation 12) is proportional to the number of particles, we can conclude that the application of the stochastic approach in systems involving a large number of particles and with only two physical particles colliding per MC cycle is highly impractical.

The procedure previously described is not very useful when simulating a cloud large volume, because of the high cost in computation. For example, in a three dimensional cloud model the typical coalescence cell has a volume of  $10^9 \text{ cm}^3$  and considering a droplet concentration at cloud base typical of maritime clouds ( $10^2 \text{ cm}^{-3}$ ), then the number of droplets will be about  $10^{11} \text{ cm}^{-3}$ . A possible solution to this problem relies in the implementation of the grouping method developed by Ormel and Spaans (2008), where particles of the same species are divided into groups and only collisions between groups of identical particles are considered in a MC cycle.

A similar approach for the stochastic collection (the Super Droplet method) was proposed by Shima et al. (2007). They defined collision between Super Droplets (that are actually species): droplets with the same attributes and position) and show that the result of Monte Carlo scheme agrees with the solution of the kinetic collection equation for the one component case. Nevertheless, for the Super Droplet method to reproduce accurately the solution of the kinetic collection equation, the number of species should be extremely large (around  $2^{17}$ ). As a consequence, the method doesn't reproduce well, for example, the time evolution of a monodisperse initial condition and will be discarded as an option to model the collection process for large coalescence volumes. In the grouping method (Ormel and Spaans, 2008) the species are divided into groups composed of identical particles. Thus, for the number of species:

$$N_i = w_i 2^{z_i} \quad (16)$$

Where  $w_i$  is number of groups, and  $z_i$  is the zoom number. Then, instead of tracking  $N_i$  droplets, we will simulate the collision between  $w_i$  groups, each containing  $2^{z_i}$  droplets. The number of groups will now determine the collision rate. The coagulation is accelerated significantly, because collisions are now between groups of particles, and not between individual particles. The total number of physical particles is:

$$N_T = \sum_{i=1}^{N_s} w_i 2^{z_i} \quad (17)$$

where  $N_s$  is the total number of species. The collision rates between groups of different species are calculated in the form (if  $z_i \leq z_j$ ):

$$a^G(i,j) = V^{-1} K(i,j) N_i N_j dt / 2^{z_i} \equiv \text{Pr} \{ \text{Probability that two groups of species } i \text{ and } j \text{ (for } i \neq j \text{) with populations (number of particles) } N_i \text{ and } N_j \text{ will collide within the imminent time interval} \} \quad (18)$$

And the collision between groups of the same species:

$$a^G(i,i) = \frac{V^{-1} K(i,i) N_i (N_i - 1)}{2} / 2^{z_i - 1} \equiv \text{Pr} \{ \text{Probability that two groups of the same species } i \text{ with population (number of particles) } N_i \text{ collide within the imminent time interval} \} \quad (19)$$

In the general case  $z_i \leq z_j$ , then each  $i$  particle collides with  $2^{z_j - z_i}$   $j$  particles. After the collision event, only one group consisting of  $2^{z_i}$  particles is obtained. For the new particles, the mass of the  $k$ -component is calculated as:

$$m_{ki} + 2^{z_j - z_i} m_{kj} \quad (20)$$

The grouping algorithm in the form implemented in Ormel and Spaans (2008) is not feasible for incorporating into a microphysical framework, because the simulation results are the averages over several realizations, and we need a single realization and a constant time step for linking to a microphysical model. Thus, the modification of the Monte Carlo algorithm proposed by Sun et al. (2007) is also implemented for the grouping method. The algorithm can be summarized as follows:

1. At  $t=0$ , the event counter is set to zero and the initial number of species  $N_1, N_2, \dots, N_N$  are defined.
2. Set the zoom numbers for each species ( $N_i = w_i 2^{z_i}$ )
3. Determine the total number of group collisions ( $C_T^G$ ) in the time interval  $\Delta t$  by using the expression (where  $a^G(i, j)$  is calculated according to Eqs. 18 and 19) :

$$C_T^G = \frac{\Delta t \sum_{i=1}^{N_s} \sum_{j=1}^{N_s} a^G(i, j)}{V} \quad (21)$$

4. Generate  $C_T^G$  random ( $r_i$ ) numbers from a uniform distribution and determine the collision indexes from the relation:

$$\sum_{v=1}^{p-1} a_v^G < r_i \alpha^G < \sum_{v=1}^p a_v^G \quad (22)$$

In (22),  $\alpha_v^G = a^G(i, j)$  and  $\alpha^G = \sum_{v=1}^{\frac{N(N+1)}{2}} a_v^G$ , where the index  $\{v\}$  defined the total collision space.

5. Reduce the number of groups for the colliding species:  $w_i = w_i - 1$ ,  $w_j = w_j - 1$ , and the number of physical particles for the species: If  $z_i \leq z_j$  then  $N_i = N_i - 2^{z_i}$  and  $N_j = N_j - 2^{z_i}$ . For the new species created in the collision the number of particles is increased by  $2^{z_j - z_i}$ .
6. Return to step 2.

The performance of the algorithm was checked again by comparison with the analytical solution for the sum kernel (Eqs.15 a, b). We have calculated the evolution of an initial monodisperse distribution of  $10^8$  droplets of  $14 \mu\text{m}$  in radius (droplet mass  $1.1494 \times 10^{-8} \text{g}$ ) in a cloud volume of  $10^6 \text{cm}^3$ . As was pointed out, only a single stochastic experiment was run with a zooming factor ( $z_i$ ) of 10. Thus, the number of particles in each group was  $2^{10} = 1024$ . Figure 4 displays the comparison between the analytical and the MC total concentrations, indicating a very good correspondence between the two methods.

An additional simulation was performed, using the constant kernel, and the results compared with the analytical solution of the two-component kinetic collection equation:

$$\begin{aligned} \frac{\partial N(m, n; t)}{\partial t} = & \frac{1}{2} \sum_{m'=0}^m \sum_{n'=0}^n K(m-m', n-n'; m', n'; t) N(m-m', n-n'; t) N(m', n'; t) \\ & - N(m, n; t) \sum_{m'=0}^{\infty} \sum_{n'=0}^{\infty} K(m, n; m', n') N(m', n'; t) \end{aligned} \quad (23)$$

In (23),  $N(m, n, t)$  is the average number of particles consisting of  $m$  and  $n$  monomers of the first and second kind respectively (with water mass from size bin  $m$  and aerosol mass from size bin  $n$ ). The water mass in size bin  $m$  equals the volume of a droplet in the smallest (monomer droplet) bin multiplied by  $m$ , the aerosol mass in size bin  $n$  equals the volume of an aerosol in the smallest bin (monomer aerosol) multiplied by  $n$ .

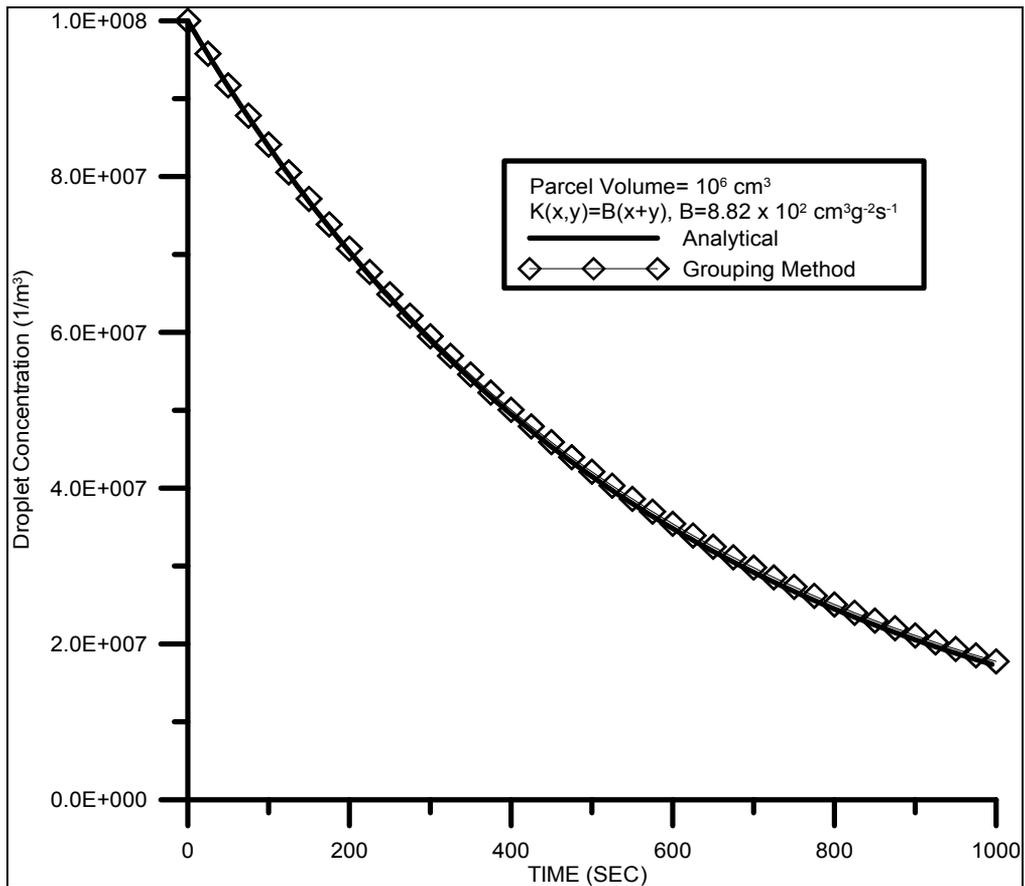


Fig. 4. Time evolution of the total number of particles obtained from the grouping method (diamonds), versus the analytical solution from the kinetic collection equation (solid line).

Solutions to (23) can be obtained for an important class of collection kernels, such as when the kernel depends only on the total number of monomers (droplets and aerosols) in each colliding particle. In this case:

$$K(m,n;m_1,n_1)=K(m+n,m_1+n_1) \quad (24)$$

Lushnikov (1975) constructed an explicit form for the composition distribution for this type of kernel, which corresponds to coagulation of initially monomeric particles. In this case  $N(1,0;0)=c_1$  and  $N(0,1;0)=c_2$ , corresponding to the situation with initially  $c_1$  droplets and  $c_2$  aerosols. The composition distribution may be expressed as (Lushnikov, 1975):

$$N(m,n;t)=\binom{m+n}{n}\left(\frac{c_1}{c_0}\right)^m\left(\frac{c_2}{c_0}\right)^n N(m+n,t) \quad c_0=c_1+c_2 \quad (25)$$

where  $\binom{m+n}{n}$  are the binomial coefficients, and  $N(m+n,t)$  is the number of particles composed of  $(m+n)$  monomers ( $m$  monomer droplets and  $n$  monomer aerosols). Lushnikov

(1975) showed that  $N(m+n,t)$ , for the type of kernels (24) is a solution of the one-component kinetic collection equation (1). For the constant kernel  $K(m,n;m_1,n_1)=A$  and a monodisperse initial distribution with concentration  $c_0$ , the analytical size distribution for the one-component KCE is:

$$N(i,t)=4c_0 \frac{(T)^{i-1}}{(T+2)^{i+1}} \quad \text{with} \quad T=Ac_0t \quad (26)$$

Then, for the constant kernel, the analytical solution of Eq. (23), calculated according to the expression (25) for the constant kernel  $K(m,n;m',n')=1.2 \times 10^{-4} (\text{cm}^3 \text{sec}^{-1})$  was compared with results of the Monte Carlo two-component simulation which was conducted for initially monomeric particles (droplets and aerosols) with concentrations  $c_1=30000$  and  $c_2=30000$  ( $N(1,0;0)=30 \times 10^3$  and  $N(0,1;0)=30 \times 10^3$ ). The initial volume was set equal to  $1000 \text{ cm}^3$ . The results are displayed in Figs. 5a,b. Again, a good agreement between the two approaches is found. These results support the validity of the grouping method for two-component stochastic coalescence.

#### 4. The multicomponent microphysical framework

The stochastic algorithm described in section 3 was incorporated into a multicomponent cloud microphysical framework. This particle-based cloud microphysical model will explicitly resolve the composition of individual droplets containing different types of CCN and is designed to accurately track the evolution by activation, condensation and coalescence of the composition of individual droplets with internally or externally mixed aerosols.

##### 4.1 Modeling of dynamical processes

The microphysical model is coupled with a simple parcel model. The air parcel is assumed to be adiabatic and homogeneous with no heat and mass exchange with the environment, with a pressure that adjusts instantaneously to that of the surrounding air, which is in hydrostatic equilibrium. The vertical velocity is prescribed. The set of equations for this case has the form (Pruppacher and Klett, 1997):

$$\frac{dT}{dt} = -\frac{gU}{c_{pa}} + \frac{L_e}{c_{pa}} C_{ph} \quad (27a)$$

$$C_w = \frac{dQ_L}{dt} \quad (27b)$$

$$\frac{dQ_V}{dt} = -C_w \quad (27c)$$

Where  $T$  is the temperature;  $U$ , the vertical velocity;  $g$ , the acceleration of gravity,  $c_{pa}$  the specific heat of air;  $L_e$ , the latent heat of evaporation;  $Q_V$  and  $Q_L$  are the water vapor and water mixing ratios and  $C_w$  is the rate of condensation. There is no sedimentation of drops with this simplified treatment of the dynamics.

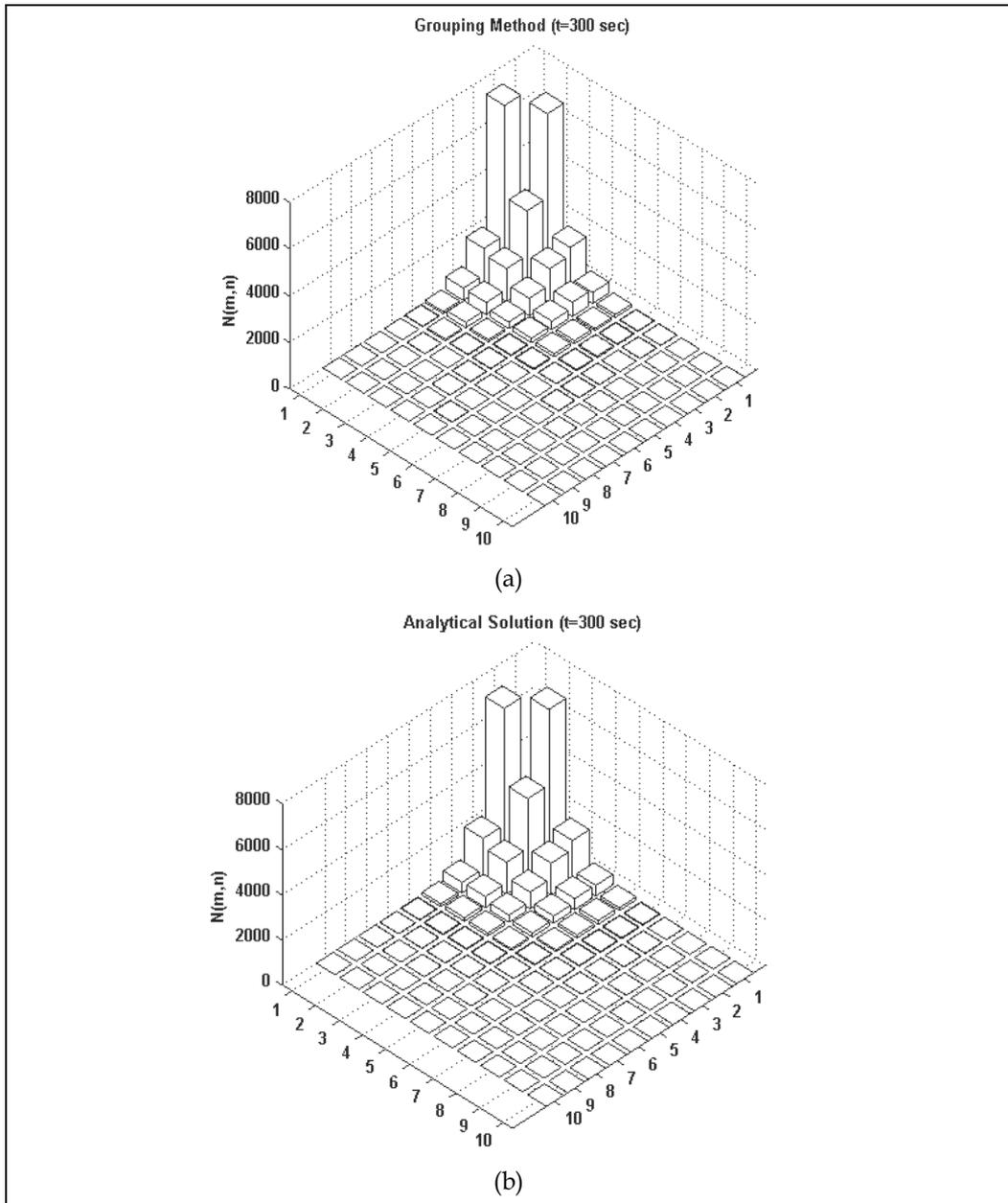


Fig. 5. Two dimensional droplet distribution  $N(m,n)$  for the constant kernel obtained by a) the grouping method and b) from the analytical solution of the two-component KCE. Simulations were conducted with initial conditions  $N(1,0)=30000$  and  $N(0,1)=30000$ .

#### 4.2 Condensation and evaporation of droplets

The usual form of the growth equation is not feasible for multicomponent microphysics. Therefore, we will consider the mass change of the species through the condensation -

evaporation process according to the modified form of the Köhler theory proposed by Mircea et al. (2002):

$$S = \frac{2\sigma M_w}{R_v T \rho_w r} - \frac{3\Phi_s M_w}{4\pi \rho_w r^3} \times \left( \sum_{i, \text{inorg}} \frac{v_i m_i}{M_i} + \sum_{j, \text{org}} \frac{v_j m_j}{M_j} \right) \quad (28)$$

This form of the Köhler equation takes into account the presence of multiple components (water soluble organic compounds, WSOC and inorganic salts) in the CCN. In (28)  $S$  is the supersaturation ratio,  $M_w$  and  $\rho_w$  are the molecular mass and density of water,  $\sigma$  is the surface tension,  $\Phi_s$  is the osmotic coefficient ( $\Phi_s=1$ ),  $R_v$  is the gas constant,  $T$  is the temperature, and  $r$  is the droplet radius. The number of dissociated ions, soluble mass and molecular mass respectively of the inorganic and organic components of CCN particles are represented by  $v_i, m_i, M_i$  and  $v_j, m_j, M_j$ . An ideal solution is assumed ( $\rho_s = \rho_w$ ).

Then, according to (28) the governing equation for diffusional growth of a water droplet of radius  $r$  is (Rogers and Yau, 1989):

$$r \frac{dr}{dt} = \frac{(S-1) \frac{2\sigma M_w}{R_v T \rho_w r} + \frac{3\Phi_s M_w}{4\pi \rho_w r^3} \times \left( \sum_{i, \text{inorg}} \frac{v_i m_i}{M_i} + \sum_{j, \text{org}} \frac{v_j m_j}{M_j} \right)}{F_k + F_d} \quad (29)$$

$$F_k = \left( \frac{L}{R_v T} - 1 \right) \frac{L \rho_w}{kT} \quad (30)$$

$$F_d = \frac{\rho_w R_v T}{De_s(T)} \quad (31)$$

Here  $S$  is the ambient, saturation ratio,  $F_k$  represents the thermodynamic term associated with heat conduction,  $F_d$  is the term associated with vapor diffusion. In (30) and (31)  $R_v$  is the individual gas constant for water vapor,  $k$  is the coefficient of thermal conductivity of air,  $D$  is the molecular diffusion coefficient,  $L$  is the latent heat of vaporization and  $e_s(T)$  is the saturation vapor pressure.

In order to allow larger integration steps for the condensation process an implicit Euler discretization scheme was adopted:

$$\frac{r_{n+1}^2 - r_n^2}{2\Delta t} = \frac{(S-1) \frac{a}{r_{n+1}} + \frac{b}{r_{n+1}^3}}{F_k + F_d} \quad (32)$$

$$\text{Where } a = \frac{2\sigma M_w}{R_v T \rho_w} \text{ and } b = \frac{3\Phi_s M_w}{4\pi \rho_w} \times \left( \sum_{i, \text{inorg}} \frac{v_i m_i}{M_i} + \sum_{j, \text{org}} \frac{v_j m_j}{M_j} \right)$$

The droplet radius  $r_{n+1}$  in the  $n+1$  iteration were calculated with the Newton Raphson method. In the model there is no need to parameterize the activation process since the equation (29) was applied to both the unactivated equilibrium droplets and activated drops. In the first case, the numerical solution of (29) gives the equilibrium radius for a given saturation ratio, which satisfies the Köhler equation:

$$S=1+\frac{a}{r_{n+1}}-\frac{b}{r_{n+1}^3} \quad (33)$$

### 4.3 Treatment of supersaturation

For calculating the saturation ratio, a time splitting procedure was used, and the evolution of the variables due to dynamical processes is calculated first:

$$T^*=T^n-\Delta t \times \frac{gU}{c_{pa}}, \quad Q_v^*=Q_v^n \quad (34)$$

There is no change in the water vapor mixing ratio due to dynamics because the air parcel is assumed to be adiabatic with no mass exchange with the environment. By taking into account the microphysical processes, the temperature and water vapor at the  $n+1$  time step are calculated as:

$$T^{n+1}=T^*+\frac{\Delta t}{c_{pe}} \times L_e \frac{d\chi}{dt}, \quad Q_v^{n+1}=Q_v^*-\Delta t \times \frac{d\chi}{dt} \quad (35)$$

Where  $\Delta t$  is the time step, and  $d\chi/dt$  is the condensation rate, which is calculated from the expression:

$$\frac{d\chi}{dt}=\frac{\rho_w}{\rho_a} \sum_i N_i 4\pi r_i^2 \frac{dr_i}{dt} \quad (36)$$

Here,  $N_i$  is the total number of droplets (unactivated and activated) for the species with index  $i$ ,  $r_i$  is the droplet radius,  $\rho_w$  and  $\rho_a$  are the water and air densities, and  $dr_i/dt$  is calculated from (29). The saturation ratio at the  $n+1$  time step can be found from the equation (Hall, 1980):

$$S^{n+1}=\frac{Q_v^{n+1}}{Q_{vs}^{n+1}(p,T^{n+1})}=\frac{Q_v^*-\Delta t \times \frac{d\chi}{dt}}{T^*+\frac{\Delta t}{c_p} \times L_e \frac{d\chi}{dt}} \quad (37)$$

Which is solved iteratively using the secant method. In (37) the condensation rate  $d\chi/dt$  is evaluated at  $\bar{S}_m=0.5(S^n+S_m^{n+1})$  and  $S_m^{n+1}$  is iteratively determined until  $|S_m^{n+1}-S_m^{n+1}|<10^{-8}$ . After that, the saturation ratio  $\bar{S}_m=0.5(S^n+S_m^{n+1})$  is applied directly to the growth equations that are integrated implicitly.

## 5. Simulation results

The parcel model described in section 4 was used to simulate the microphysical evolution of an orographic cloud sampled on November 20, 2007, located in the northwest corner of Nebraska.

The initial CCN distribution for the simulation represented an external mixture of 3 different composition categories: pure sodium chloride (NaCl), oxalic acid-elemental carbon

mixture (OC-EC) and ammonium sulfate-elemental carbon:  $((\text{NH}_4)_2\text{SO}_4\text{-EC})$  (Figure 6). The initial CCN total number concentration was  $531 \text{ cm}^{-3}$ . The (OC-EC) and  $((\text{NH}_4)_2\text{SO}_4\text{-EC})$  particles were assumed to consist of 90% water soluble materials, with a 10% of elemental carbon. NaCl particles were assumed 100% solubility. The smallest NaCl aerosol particles require a supersaturation of 0.33 % to activate, a value that was never exceeded in the simulations (see Figure 7b). The largest critical supersaturation for the OC-EC particles is 0.24 %.

The air parcel ascends with a constant vertical velocity of  $0.5 \text{ ms}^{-1}$  from cloud base at  $-8^\circ\text{C}$  and 764 hPa. The calculation starts at 98% relative humidity with moist adiabatic lapse rate. A cloud height from the cloud base of 200 m was simulated. The parcel volume was assumed to be  $10^5 \text{ cm}^3$  (100 liters), with zoom numbers for the grouping algorithm  $z_i=5$  for all the species (which means we have  $2^5=32$  droplets in each group), and a time step of  $\Delta t=1$  sec. Therefore, the initial number of particles in the simulation was  $531 \times 10^5$ .

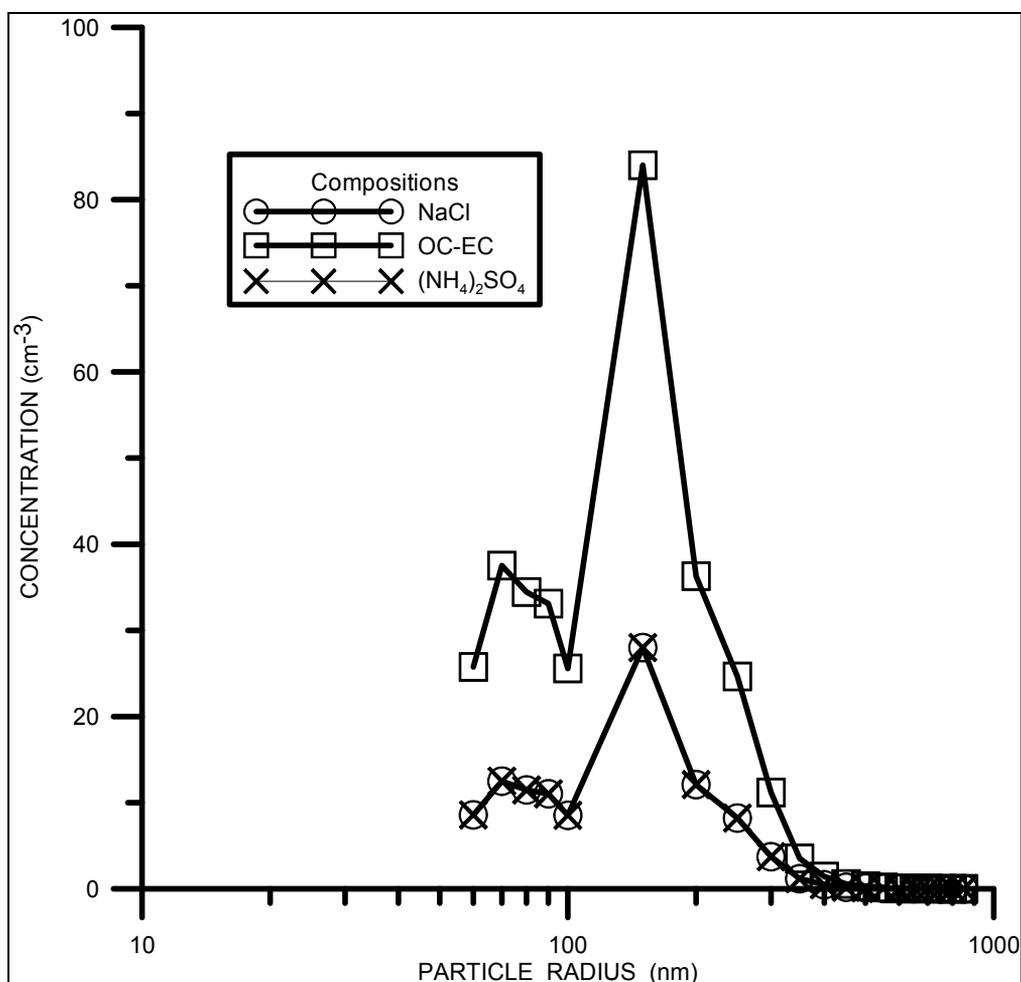


Fig. 6. CCN size distributions that served as basis of calculations for the three different compositions.

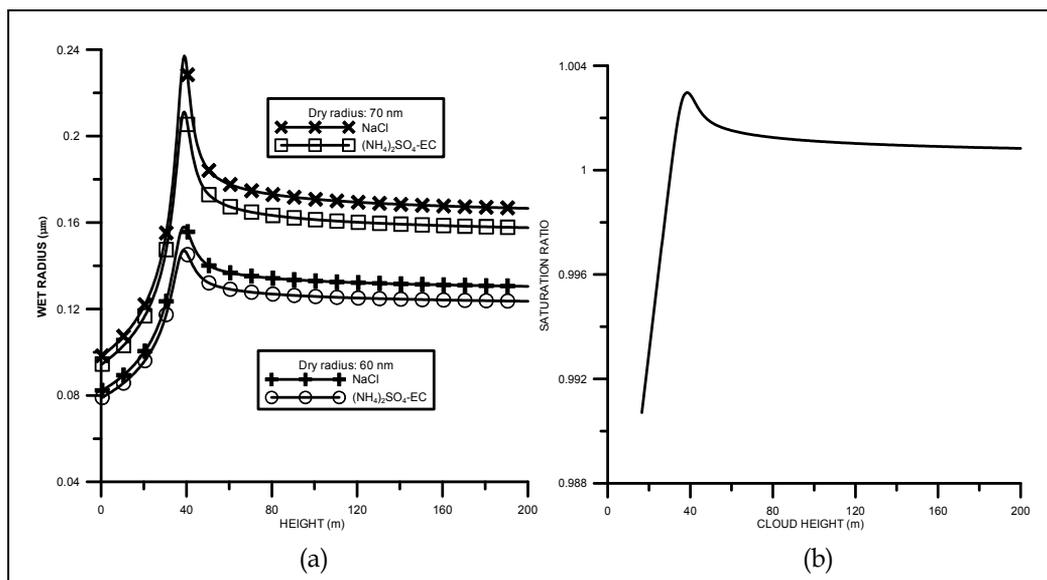


Fig. 7. Wet radius for interstitial aerosols as a function of height above cloud base, and supersaturation profile for the simulated cloud.

In a first experiment, the size distribution was allowed to evolve only by nucleation and condensation. The vertical profile of supersaturation obtained in this simulation is shown in Figure 7b. As can be observed, the maximum supersaturation is about 0.3%.

After peak supersaturation is reached, the aerosol number total concentration decreased from  $531 \text{ cm}^{-3}$  to  $42 \text{ cm}^{-3}$  due to nucleation scavenging (a 92% decrease), in agreement with field observations of Hegg & Hobbs (1983). The number concentration of the different components show the following evolution: the NaCl and (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>-EC aerosol particles decreased by 80% (from 106 to 21  $\text{cm}^{-3}$ ) and 100% of the OC-EC aerosol particles were nucleated. This is consistent with the cloud supersaturation spectra (CSS) which have maximum critical supersaturations of 0.25% for the aerosol particles with OC-EC compositions. Consequently, all the particles in this category get activated after the maximum supersaturation of 0.3% is reached. The aerosol particles that were not activated to droplets remain as interstitial aerosols and in equilibrium with ambient supersaturation conditions (Figure 6a).

Figure 8 shows the evolution as a function of height above cloud base for the largest droplet in each of the three aerosol composition species. The droplets formed on the (OC-EC) CCN achieve a larger size than the droplets that contain the inorganic salts.

In a second simulation, the size distribution was allowed to evolve by activation, condensation and the collision-coalescence process in the manner described in section 4. The zoom numbers were assumed to be  $z_i=5$  for all the species (which means we have  $2^5=32$  droplets in each group), with a time step of  $\Delta t=1$  sec. Despite the large number of particles for the parcel under analysis, the coalescence was almost negligible with a maximum of 3 collision events per time step. Nevertheless, at the end of the parcel ascent the number of species was incremented up to 593. The huge increment in the number of species (from 57 to 593) is only explained by the coalescence, since the condensation and activation processes

conserved the number of species. Due to the fact that we have a four-component system (droplet radius and three types of aerosols inside droplets), practically every collision leads to the formation of a new species. The maximum droplet radius was  $7.35\ \mu\text{m}$  with an aerosol with composition OC-EC and radius  $0.3999\ \mu\text{m}$ . An additional simulation was performed with a parcel volume of  $5000\ \text{cm}^3$  and a zoom factor of  $z=0$  for all the species in order to compare with the previous simulation. In that case we were considering collisions between particles, not between groups. The same maximum droplet radius was obtained as a result of the simulation. These preliminary findings are encouraging and show the potential of this modeling approach.

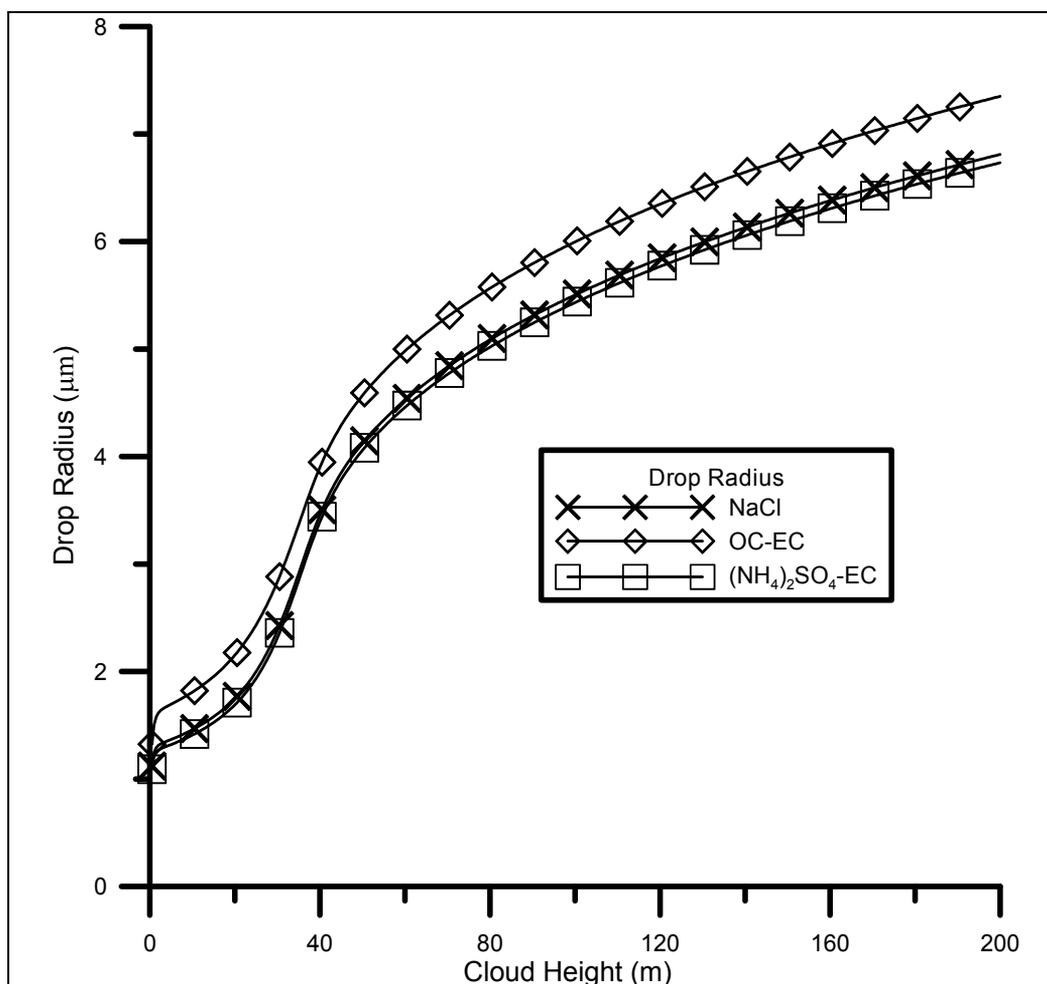


Fig. 8. Largest droplet radius for the activated species.

## 6. Conclusions

In this work, a novel Monte Carlo multicomponent framework for the collection process was introduced, and its characteristics discussed in detailed. The Monte Carlo algorithm is based on the grouping method proposed by Ormel and Spaans (2008), and allows accurate simulation of the coalescence process in large cloud volumes with reasonable cost in computation. Therefore, it can be a useful tool to simulate microphysical evolution in cloud models with complex dynamics.

The applicability of the Monte Carlo grouping method was demonstrated by linking the stochastic framework with a microphysical model with simple dynamics, and presenting very preliminary results of an orographic cloud formation with four component microphysics.

Simulation results suggest that the Monte Carlo grouping method can be computationally more efficient than the deterministic framework (based on the solution of the KCE), when the number of components of the system is larger than 2. Then, it is expected to be more feasible for modeling complicated microphysics, and provides us with a new tool to solve open problems in cloud modeling.

Even though a more thorough validation of the method is still needed, we believe that this stochastic algorithm will prove to be a useful new approach to simulations of multicomponent microphysics. It is particularly applicable to studies of cloud and aerosol interactions with multiple types of CCN, the modeling of collection process in mixed phase clouds and cloud chemistry.

As a future work, an important effort will be required to extent this method to model the microphysical evolution of mixed phase clouds, and to include the chemical processes. This work attempts to be a first step toward the accomplishment of these goals.

## 7. Acknowledgements

The authors are grateful to LUFAC Computación SA de CV for funding the publication of this work.

## 8. References

- Alfonso, L. & Raga, G.B. (2004). The influence of organic compounds in the development of precipitation acidity in maritime clouds. *Atmos. Chem. Phys.*, 4, 1097-1111.
- Alfonso, L.; Raga G.B., & Baumgardner, D. (2008). The validity of the kinetic collection equation revisited.-Part II.: Simulations for the hydrodynamic kernel. *Atmos. Chem. Phys.*, 8, 969-982.
- Alfonso, L.; Raga, G.B. & Baumgardner, D. (2009). Monte Carlo simulations of two component droplet growth by stochastic coalescence. *Atmos. Chem. Phys.*, 9, 2141-1251.
- Alfonso, L.; Raga G.B., & Baumgardner, D. (2010). The validity of the kinetic collection equation revisited.-Part II.: Simulations for the hydrodynamic kernel. *Atmos. Chem. Phys.*, 10, 6219-6240.
- Bayewitz, M.H.;Ye Yerushalmi; J., Katz, S. & Shinnar, R. (1974). The extent of correlations in a stochastic coalescence process, *J. Atmos. Sci.*, 31, 1604-1614.

- Berry E.X. & Reinhardt R.L. (1974). An analysis of cloud drop growth by collection. Part I. Double distributions. *J. Atmos. Sci.*, 1974, 31, 1814-1824.
- Bott, A.A. (2000). A flux method for the numerical solution of the stochastic collection equation: Extension to two-dimensional particle distribution. *J. Atmos. Sci.*, 57, 284-294.
- Bott, A.A. (1998). A flux method for the numerical solution of the stochastic collection equation. *J. Atmos. Sci.*, 55, 2284-2293.
- Drake, R.L. (1972). The scalar transport equation of coalescence theory: Moments and kernels, *J. Atmos. Sci.*, 29, 537-547.
- Flossmann, A. I. (1994). A 2-D spectral model simulation of the scavenging of gaseous and particulate sulfate by a warm marine cloud. *Atmos. Res.*, 32, 233-248.
- Feingold, G. & Kreidenweiss, S.M. (2002). Cloud processing of aerosol as modeled by a large eddy simulation with coupled microphysics and chemistry. *J. Geophys. Res.*, 107, 4687.
- Gillespie, D. T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, 22, 403-434.
- Hegg, D.A. & Hobbs, P.V. (1983). Preliminary measurements on the scavenging of sulfate and nitrate by clouds. *Precipitation scavenging, dry deposition and resuspension. Vol. I*, Elsevier Science, 78-89.
- Inaba, S. ; Tanaka, H.; Ohtsuki, K. & Nakazawa, K. (1999). High-accuracy statistical simulation of planetary accretion: I. Test of the accuracy by comparison with the solution to the stochastic coagulation equation, *Earth Planet Space*, 51, 205-217.
- Jacobson, M.Z. (1999). *Fundamentals of atmospheric modeling*. Cambridge University Press, 656 pp., ISBN 0521-63717.
- Khain, A.P. & Sednev, I., (1995). Simulation of hydrometeor size spectra evolution by water-water, ice-water and ice-ice interactions. *Atmos. Res.*, 36, 107-138.
- Long, A.B. (1974). Solutions to the droplet collection equation for polynomial collection kernels, *J. Atmos. Sci.*, 31, 1040-1051.
- Lushnikov, A.A. (1975). Evolution of coagulating systems III: Coagulating mixtures, *J. Coll. Int. Sci.*, 54, 94-101.
- Mircea, M.; Facchini, M.M.; Decesari, S.; Fuzzi, S. & Charlson, R.J. (2002). The influence of organic aerosol component on CCN supersaturation spectra for different aerosol types. *Tellus*. 54B, 74-81.
- Ormel, C.W. & Spaans, M. (2008). Monte Carlo simulation of particle interactions at high dynamic range: Advancing beyond the Googol. *ApJ.*, 684, 1291.
- Pruppacher, H.R. & Klett, J.D. (1997). *Microphysics of clouds and precipitation*, Kluwer Academic Publishers, ISBN 0-7923-4211-9.
- Rogers, R.R. & Yau, M.K. (1989). *A short course in cloud physics*, Elsevier, New, York.
- Shima, S.I.; Kusano, K.; Kawano, A.; Sugiyama, T. & Kawahara, S. (2005). Super-Droplet method for the numerical simulation of Clouds and Precipitation: A particle-based microphysics model coupled with non-hydrostatic model. *Arxiv:physics/0701103v1*.

- Sun, Z.; Axelbaum, R. & Huertas, J. (2007). Monte Carlo simulation of multicomponent aerosols undergoing simultaneous coagulation and condensation, *J. Aer. Sci. Tech.*, 38, 963-971.
- Scott, W.T. (1968). Analytic studies of cloud droplet coalescence, *J. Atmos. Sci.*, 25, 54-65.
- Tanaka, H. & Nakazawa, K. (1994). Validity of the statistical coagulation equation and runaway growth of protoplanets, *Icarus*, 107, 404-412.
- Wang, L.P.; Xue, Y.; Ayala, O. & Grabowski, W.W. (2006). Effect of stochastic coalescence and air turbulence on the size distribution of cloud droplets, *Atmos. Res.*, 82, 416-432.

# Monte Carlo Simulation of Room Temperature Ballistic Nanodevices

Ignacio Íñiguez-de-la-Torre, Tomás González,  
Helena Rodilla, Beatriz G. Vasallo and Javier Mateos  
*Universidad de Salamanca, Departamento de Física Aplicada  
Spain*

## 1. Introduction

The widespread use of digital broadband communications generates a huge amount of data to be processed and transmitted in the fastest possible way. To this end, the development of new electronic devices, digital and analog, able to perform data processing at ultra-high bit rates and to transmit at high frequency is necessary. In the last decade, higher frequencies have been obtained when downscaling traditional HEMTs with high In content InGaAs channels providing cut-off frequencies above 1 THz [Lai et al., 2007]. However, the reduction of the gate length below 30 nm does not provide an improvement of the device performance, obstructed by the influence of the device parasitics and the appearance of important short channel effects. A wide variety of alternatives have been proposed in order to continue improving high frequency performance of electronic devices. One promising solution is the use of new nanometer-scale ballistic transport devices based on high mobility III-V compound heterostructures. The use of advanced electron-beam lithography tools and conventional epitaxial growth techniques for III-V materials also allows the fabrication of a two-dimensional electron gas (2DEG) structures with sizes smaller than the electron mean free path ( $l_m$ ). In these new devices, by smartly custom-made geometries, electrons move like billiard balls, guided by strategically placed shapes, edges and internal deflectors rather than applied voltages [Song, 2004]. The nano-scale dimension of these devices facilitates electron transport with almost no scattering event and is referred as ballistic transport or at least quasi-ballistic and can be observed even at room temperature (RT). This involves not only the presence of velocity overshoot, which can improve the cut-off frequency of classical FET devices, but also gives birth to some new effects and applications with specially designed structures. The great performance of these nanodevices obtained with InGaAs in the channel can be even better, providing higher cut-off frequencies, by using narrow bandgap semiconductors like InAs and InSb [Suyatin et al., 2008].

Now, we introduce some examples of this *family* of nanostructures. In particular, Three-Branch Junctions [Fig. 1(a)] (TBJs, T- and Y-shaped) [Palm & Thylen, 1992; Palm & Thylen, 1996; Worschech et al., 2001a; Shorubalko et al., 2001; Xu, 2001] exhibit a parabolic negative potential at the central branch when biasing in push-pull fashion the left and right branches (in contrast with the zero central potential expected for diffusive structures) [Mateos et al., 2003a; Mateos et al., 2004; Rashmi et al., 2005; Bednarz et al., 2006; Íñiguez-de-la-Torre et al., 2007]. Similar effects take place in ballistic rectifying devices achieved by inserting an

obstacle (antidot) of triangular or diamond shape in the centre of a ballistic cross junction [Song et al., 1998; Song et al., 2001, Vasallo et al., 2004; González et al., 2004]. This type of geometry has also recently been improved by attaching two strategically placed in-plane gates for the fabrication of the so called Ballistic Deflection Transistors (BDT) [Wesström, 1999; Hieke & Ulfward 2000; Diduck et al., 2009; Kaushal et al., 2010]. Ultrafast rectifying nano-diodes, named Self-Switching Diodes (SSDs) [Fig. 1(b)], made with a single-step lithographic process that breaks the symmetry of a narrow channel, have also been successfully fabricated [Song et al., 2003] and simulated [Mateos et al., 2005]. As a consequence of their nonlinear properties, all of these devices have demonstrated manifold functionalities as: wave rectification, frequency doubling and Boolean logic operation, and can work at very high frequencies. Perhaps, most important from the point of view of circuit applications, is that they work at RT. For example, SSDs have been shown to operate as detectors even reaching the THz range [Balocco et al., 2005; Balocco et al., 2008]. However, due to its large impedance and the influence of contact parasitic capacitances, mixing and frequency doubling in TBJs at RT have only been measured below the GHz range [Lewén et al., 2002], and the rectifying effect up to 20 GHz [Worschech et al., 2001b; Worschech et al., 2002] (up to 40 GHz in a double Y branch configuration [Bednarz et al., 2005]). This new *family* of ballistic nanodevices are ideal candidates to constitute the base for the fabrication of analog/digital circuits to process and transmit data at THz frequencies and at RT. Moreover, these material systems ensure a full compatibility with HEMT technology, which can be used for the post-processing of the output signals.

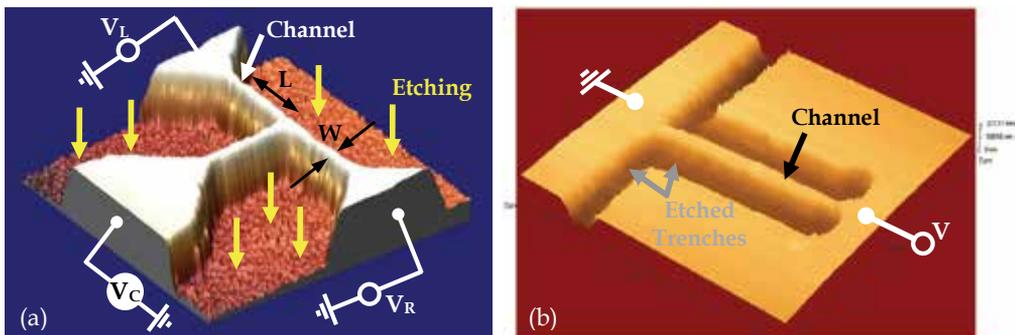


Fig. 1. (a) Atomic force micrographs of a typical (a) TBJ and (b) SSD.

When such new and promising strategies are explored, the test-and-error approach is not efficient at all. At this point, the development of theoretical models and numerical simulations can be of great help. Under these ballistic or quasiballistic conditions, the classic drift-diffusion or hydrodynamic models, traditionally used for device simulation and design, are not appropriate any more. Even if some theoretical descriptions of the operation of ballistic devices have been proposed, always starting from a coherent transport model based on the Landauer-Büttiker formalism [Landauer, 1957; Büttiker, 1986; Song, 1999], the most adequate numerical technique is the Monte Carlo (MC) method [Jacoboni & Lugli, 1989], especially for RT operation. In particular, the MC technique incorporates in a natural way all the scattering mechanisms, so it is able to correctly account for ballistic or quasi-ballistic transport and provide not only static results but also the dynamic and noise behaviour of the devices. Monte Carlo simulations have been proven as an exceptional useful tool for the optimization of the ballistic nanodevices like cross junctions [González et al., 2004], BDTs [Kaushal et al., 2010],

or the TBJ-based multiplexor/de-multiplexor [Mateos et al., 2003a]. In this chapter, by means of our Monte Carlo simulators, we will focus our attention on designing and optimizing TBJs and SSDs to operate at sub-millimetre frequencies and RT. Taking as a base the knowledge of the internal microscopic processes, we will exploit the ballistic effects when the size of the structures is reduced below the electron mean free path. Monte Carlo simulations are able to identify and explain the physical origin of the nonlinear effects in this *family* of nanostructures: (i) electrostatic effects typically associated with the presence of surface charges and (ii) asymmetrical charge distributions related to ballistic transport. Due to the nanometer length associated with ballistic transport, intrinsic operation up to THz frequencies is expected [Mateos et al., 2003b]. However, to really exploit the estimated intrinsic high speed these technologies require more efforts directed towards a proper design of accesses. Reduce cross-talk (extrinsic) capacitances and extrinsic resistances or fabricating devices in parallel in order to reduce the intrinsic impedance without increasing the capacitance will be one of the main challenges. In addition, as the size of electronic devices is reduced, the surface/volume ratio considerably increases and, in strong contrast to conventional devices, when sizes reach the nanometer scale, surface effects can get to have a remarkable importance on electron transport, even becoming decisive in the device behaviour. Also, at this nanometer scale, modelling of the contacts has proven to be crucial. The specific models implemented in the Monte Carlo simulators to deal with these devices as dynamic surface charge models [Íñiguez-de-la-Torre et al., 2007], injection statistics [González et al., 1999], etc, will be explained in this chapter.

The outline of this chapter is as follows. Firstly, in Section II the physical model used for the MC simulation of the ballistic structures is explained. We put special emphasis on the modelling of the surface charge, providing the details of the algorithm. Section III is dedicated to examine the dependence of the well-known parabolic behaviour of the output voltage of TBJs on the size of their branches. Results concerning to the frequency response are also discussed. In Section IV we successfully apply our model to explain the physics underlying the SSD rectifying behaviour and to analyze the AC response and noise spectra dependence on the topology of the devices. A systematic study of channel length and width of the trenches is shown to provide design indications to improve their performance. Finally the main conclusions of the present work are summarized.

## 2. Monte Carlo method. Methodology

In this chapter, the ballistic nanodevices have been studied by analyzing the results obtained using a semi-classical *ensemble* MC simulator self-consistently coupled with a 2D Poisson solver (PS) [Jacoboni & Lugli, 1989]. The transport model locally takes into account the effect of degeneracy and electron heating by using the rejection technique and the self-consistent calculation of the local electronic temperature and Fermi level [Mateos et al., 2000a]. The surface charges appearing at the boundaries of the semiconductors in contact with dielectrics are also considered in the model [Mateos et al., 1999]. The validity of this simulation model has been checked in previous works by means of the comparison with experimental results of static characteristics, small signal behaviour and noise performance of a 0.1  $\mu\text{m}$  gate AlInAs/InGaAs lattice matched HEMT (InP-based) [Mateos et al., 2000b].

### 2.1 Generalities

The basic philosophy of the single particle Monte Carlo technique, applied to charge transport in semiconductors, consist of simulating the motion of a charged carrier inside the

crystal. It is intended to study the free flight of the particle accelerated by an applied electric field between instantaneous random scattering events. Describing in more detail [Fig. 2(a)], the algorithm generates random free flight times for each particle, determines the state after each free flight, randomly chooses from amongst the different scattering mechanisms at the end of the free flight, computes the final energy and momentum of the particle after scattering, and finally reiterates the routine for the subsequent free flight. By monitoring the particle motion during the simulation, it is possible to statistically estimate the magnitude of several physical parameters for the particle such as the distribution function, average drift velocity, average energy, etc.

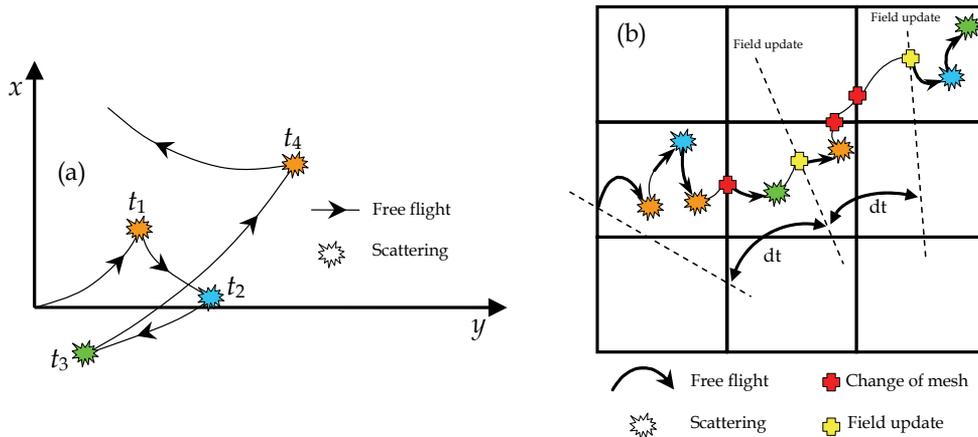


Fig. 2. (a) Simple diagram of the particle motion in the real space under a uniform electric field applied in the  $x$  direction. (b) 2D device simulator scheme.

In order to study transient behaviour, a synchronous simulation of a reasonable number of particles is indispensable. This is called *ensemble* Monte Carlo, in which the above algorithm is repeated for each particle. Every given time step in which, the individual carriers are simulated independent of the others, the quantities of interest are sampled and averaged. However, within a real semiconductor device, it is also essential to consider the internal electric potential obtained from the solution of the Poisson equation, as this is also an accelerating source for the particles. Consequently, it becomes necessary to couple both, the transport kernel, and the field solver to each other. For this purpose, a spatial grid is needed to solve the Poisson equation. In this frame, the simulation of the particle-based *ensemble* is carried out over a reasonably small time step, under the action of a self-consistent electric field (solution of Poisson equation) with the appropriate boundary conditions. At the end of each time interval, Poisson equation for the next time-step is solved again using the configuration of charges obtained from the *ensemble* Monte Carlo [Fig. 2(b)].

The electric field is computed (neglecting the inductive magnetic effects) by LU decomposition of the Poisson equation in a finite differences approach. Accurate scattering probabilities models for ionized impurity, alloy, polar and non-polar optical phonon, acoustic phonon and intervalley scattering ( $\Gamma$ -L-X) with a non-parabolic spherical valleys and effective mass approximation are used. As the size of our devices is larger than the de-Broglie electron wave length, quantum mechanical non-local effects are not taken into account. Neumann boundary conditions (the difference between the normal components of

the respective electric displacement vectors must be equal to any surface charge) are imposed in the semiconductor/dielectric boundaries, so that current only flows in/out of the device through the contacts, in which a Dirichlet condition (the potential is fixed) is imposed.

Concerning the analysis of noise, in the simulation we follow the standard scheme. The instantaneous current is calculated using the generalized Ramo-Shockley theorem [Kim et al., 1991], which evaluates the simultaneous contribution of all particles involved in the MC simulation to the total electrode current. The mathematical quantity employed for the characterization of noise is the autocorrelation function of current fluctuations. Then, by the Wiener-Kintchine theorem, the autocorrelation function is related to the noise spectra.

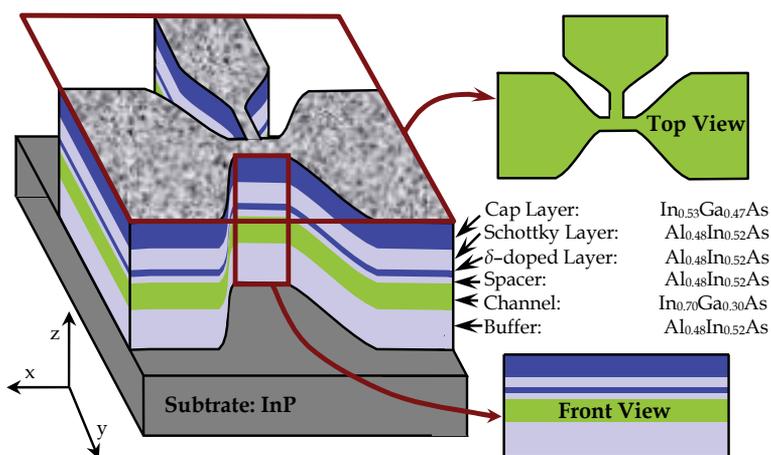


Fig. 3. Three-dimensional (3D) geometry and layer structure of a T-branch junction and scheme of the 2-D front-view (FV) and top view (TV) Monte Carlo simulations.

## 2.2 Simulations of channels: Three dimensional (3D) to two dimensional (2D) approach

For the correct modelling of nanodevices, a 3D simulation would be necessary in order to take into account the effect of the lateral surface charges and the real geometry of the structures. However, for the moment, only a 2D MC model has been developed, and some simplifications and assumptions must be made [Mateos et al., 2003a; Mateos et al., 2003b]. To account for the top geometry of the devices (for TBJs, YBJs, or ballistic diodes in our case), top-view (TV) simulations will be carried out [see Fig. 3]. They are performed in the  $xy$  plane; therefore, the real layer structure is not included, and only the channel will be simulated. In order to account for the fixed positive charges of the whole layer structure, a net doping  $N_{db}$  is assigned to the channel in TV simulations, but impurity scattering is switched off. In this way, the electron transport through the undoped channel is well reproduced, since this is a “virtual” doping  $N_{db}$  associated with the charges of the cap and  $\delta$ -doped layers. On the other hand, a negative surface charge density  $\sigma$  is assigned to the semiconductor-air interfaces to account for the influence of the surface states originated by the etching processes. The non-simulated dimension  $Z$  (used for the comparison of the simulated values of current with those measured in real devices) was estimated as  $Z = n_s / N_{db}$ , with  $n_s$  the value of sheet electron density in the fabricated channels. In addition, our approach has recently been further validated by the results of [Sadi et al., 2009]. Using a

3D model they have almost perfectly replicated the results obtained with our 2D model for the same set of TBJ junctions. Typically, InGaAs is the high-speed material used for the channel in the InP based heterojunctions. However, two different narrow band gap semiconductors, InAs and InSb, and their associated heterostructures, AlSb/InAs and AlInSb/InSb provide much higher values of mobility associated to their very small electron effective masses in the bottom of  $\Gamma$  valley (0.023 and 0.014 respectively). Our tool has been also properly adapted for these two high mobility semiconductors materials by carefully adjusting the simulation parameters in the single particle MC simulator. Experimental bulk mobilities have been reproduced by simulations:  $\mu=28000$  cm<sup>2</sup>/Vs for InAs and  $\mu=67000$  cm<sup>2</sup>/Vs for InSb [Rodilla et al., 2009].

### 2.3 Injection and physical model for the contacts

Since contact injection is a critical point when dealing with ballistic transport, the velocity distribution and time statistics of injected carriers will be accurately modelled [González & Pardo, 1996; González et al., 1999].

To compute the velocity distribution for the injected carrier,  $f_{iny}(\mathbf{v})$ , since the injected carriers are crossing the boundary between the contact and the adjacent cell inside the device, the thermal distribution,  $f_{th}(\mathbf{v})$ , should be weighted by the perpendicular velocity  $f_{iny}(\mathbf{v}) = \mathbf{v} \cdot f_{th}(\mathbf{v})$ . In this way, we account for the higher probability of particles with a large velocity to enter the device. In general the thermal distribution is the Fermi-Dirac (FD), however for non-degenerate material can be replaced by the Maxwell-Boltzmann one. A rejection method is used in case of FD because is not analytically integrable.

As concerns the time statistics, we first need to compute the number of carriers per unit of time entering in the device; which is the injection rate  $\Gamma$ . For a degenerate reservoir the injection statistics is a binomial distribution. However, for a non-degenerate reservoir, it is possible to use global Poissonian statistics in which the time between two consecutive electron injections,  $t_i$ , is generated with a probability per unit time  $P(t) = \Gamma e^{-\Gamma t}$ .

### 2.4 Surface charge modelling

The surface-to-volume ratio in nanoelectronic devices increases as the geometries are scaled down, so that, the device behaviour is more and more affected by the physical properties of the surfaces. Sidewall surface charge provokes the depletion of part of the conducting semiconductor channel as a consequence of coulombian repulsion and thus lowers the carrier density near the interface with the dielectric. In the total depletion approximation, the depletion width originated by a surface charge  $\sigma$  is  $W_d = \sigma / N_{db}$  at each side of the channel. Therefore, the effective conduction width is  $W_{eff} = W - 2W_d$ , with  $W$  the total width of the channel [Fig. 4(a)]. With the aim of extracting the experimental lateral depletion width  $W_d$ , the electrical characterization of channels with different length and width has been made. A value of  $W_d$  about 40 nm ( $\pm 10$  nm) for In<sub>0.7</sub>Ga<sub>0.3</sub>As channels [Galoo et al., 2004], corresponding in MC to a surface charge density of  $\sigma/q = (0.4 \pm 0.1) \times 10^{12}$  cm<sup>-2</sup> (using  $N_{db} = 10^{17}$  cm<sup>-3</sup>), has been obtained near equilibrium conditions.

A simple way to include the influence of this surface charge in MC simulations is to consider a model in which  $\sigma$  is fixed to the experimentally-extracted equilibrium value, and kept constant independently of the topology of the structure, position along the interface, bias and time. We will call this model as constant surface charge model. The surface charge is included as a Neumann boundary condition for the Poisson equation,  $\epsilon_2 E_2^n - \epsilon_1 E_1^n = \sigma$ ,

with  $\epsilon_i$  the permittivity and  $E_i^n$  the normal electric field in the  $i$ -th material. The applicability of this model becomes doubtful when the semiconductor becomes totally depleted (for  $W$  lower than 80 nm, so that  $W_{\text{eff}}$  becomes negative). Indeed, the physical origin of the surface charges is the trapping of electrons in surface states (located in the middle of the gap), but if the region near the surface is completely depleted, no electron would be able to reach the surface and the surface charge should decrease. In such a case ( $W_{\text{eff}} < 0$ ), if this model is used, the background doping  $N_{\text{db}}$  can not compensate the negative surface charge, charge neutrality is not ensured and unphysical high negative potentials are obtained in the simulation, providing incorrect results.

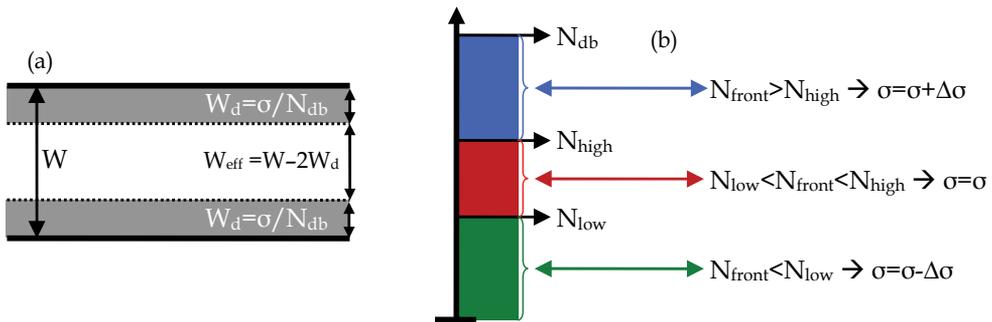


Fig. 4. (a) Effective channel width due to lateral depletion. (b) Surface charge self-consistent model.

Some features of the surface charges make difficult the possibility to implement them precisely in a MC simulator. The occupation of the surface states depends not only on its energy level but also on the potential profile and the Fermi energy in the surrounding region. Moreover, surface physical and chemical properties, fabrication processes, surface oxidation, composition and roughness determine the properties of the surface states and, as a consequence, the responses of nanometer scale devices. In addition, the capture and emission mean times of surface states (with values typically in the  $\mu\text{s}$  range) are much higher than scattering times, thus preventing their detailed treatment in a microscopic MC scheme, since a huge CPU-time would be necessary to take into account the correct dynamics of these states. For all these reasons, we have developed a new model [Íñiguez-de-la-Torre et al., 2007], based on the depletion induced by traps and not on their statistics, in which the local value of the surface charge is updated self-consistently with the carrier dynamics near the interface during the simulation. This model so called self-consistent charge model completes and improves the previous works where a constant surface charge density (neither depending on the position nor on the applied potential) was considered at the semiconductor-dielectric interfaces associated with the presence of surface states.

The philosophy of our new self-consistent model is based on the adaptation of the value of the surface charge to the carrier density in the nearby region [Fig. 4(b)]. First, we evaluate the carrier concentration next to the boundary ( $N_{\text{front}}$ ) as an average over a given number of iterations  $N_i$ . Then, it is checked if  $N_{\text{front}}$  has a value in the range  $[N_{\text{low}}, N_{\text{high}}]$  which represent the limits to which we try to adapt the electron concentration next to the interface. If the concentration ( $N_{\text{front}}$ ) is higher than the upper limit ( $N_{\text{high}}$ ), we increase the surface charge in a given amount  $\Delta\sigma$ , so that its repulsive effect provokes stronger channel depletion

and thus the concentration should diminish. On the other hand if  $N_{\text{front}}$  is smaller than the lower limit ( $N_{\text{low}}$ ), the surface charge is decreased in the same density  $\Delta\sigma$  to reduce the (too large) induced depletion. The choice of the limits  $N_{\text{low}}$ ,  $N_{\text{high}}$  is constrained by the level of statistical resolution achievable by the simulation (which depends on the number of simulated electrons). If the lower limit is too low (the forced depletion level is too strong), below the value of electron concentration that MC simulations can reliably estimate, wrong results would be obtained. More details of the values of the parameters can be found in [Iníguez-de-la-Torre et al., 2007].

To conclude, it is important to remark that evidently this “ad-hoc” surface charge model is not able to reproduce the statistics of occupation of surface states, but it does describe correctly the global effect of the surface charge. We will apply this self-consistent model to study T-shaped three-terminal ballistic junctions (TBJs) based on InAlAs/InGaAs layers and compare the results with experimental measurements, achieving a satisfactory agreement.

### 3. Three-branch junctions

The device structure of the Three-Branch Junction (TBJ) is very simple. The TBJ is a three-terminal device consisting of a T-shape (or Y-shape) conductor with three contacts at the end of each branch. The conductor is typically made of a 2DEG formed in a modulation-doped heterostructure wafer. Employing high resolution resins, electron-beam lithography and dry or wet etching, the 2DEG wafer is patterned into a T-or Y-shape structure with a dimension comparable to the mean free path ( $l_m \sim 100\text{-}300\text{ nm}$ ) at RT [Cappy et al., 2005].

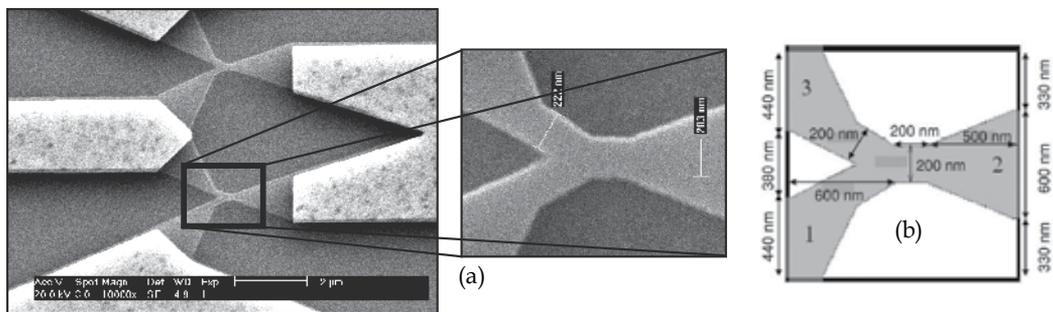


Fig. 5. (a) SEM image of a double Y-Branch junction topology with coplanar waveguide (CPW) accesses. (b) Top-view geometry (sizes correspond to the real YBJ) of the single YBJ used in the MC simulations.

Figure 5(a) shows a scanning electron microscopy (SEM) image of a typical device structure. Bright regions and dark regions indicate 2DEG layer and etched away material respectively. We also see the accesses connecting the Y-shape structure with the ohmic contacts and metal interconnects pads for DC probing (or are a part of planar transmission lines for high-frequency measurements [Irie et al., 2010]). Figure 5(b) is a screenshot of our MC tool.

#### 3.1 Overview and working principle

TBJs operation is described as follows. Input voltages are applied to the L- and R-terminals, while the C- terminal is used as the output terminal. In big devices, where  $l_m$  is smaller than the distance between contacts, each branch can be represented by a resistance [Fig. 6(a)]. So,

if we apply two separate potentials with opposite sign to the right and left terminals (push-pull fashion), the potential in the middle of the junction is zero. In this ohmic transport regime, the potential measured at the bottom of the open-circuited central branch (stem),  $V_C$ , is also zero [Fig. 6(c)]. However, when the lengths of the branches,  $L$ , are smaller than  $l_m$ , transport can not be anymore considered diffusive, but ballistic or quasiballistic. In this regime, the horizontal branches can be understood as non-linear resistances  $R(V)$ , Fig. 6(b), and; as shown by the experiments [Shorubalko et al., 2001] and theory [Xu, 2001],  $V_C$  is always negative and presents a quadratic down-bending shape  $V_C = -\alpha V^2$ , Fig. 6(c). The non-linear effect can be enhanced by decreasing the temperature of operation because the reduction in the number of scattering mechanisms [Irie et al., 2008].

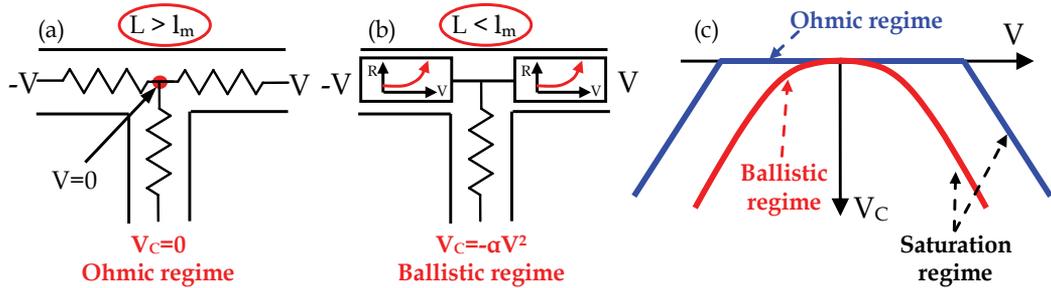


Fig. 6. Sketch of the (a) ohmic and (b) ballistic regimes of transport in TBJs and the (c)  $V_C$  response for both types of transport.

The negative values of  $V_C$  have been reproduced by Monte Carlo simulations of TBJs [Fig. 7(a)]. The explanation is related to space-charge effects originated by the joint action of (i) the surface charge at the semiconductor-air interfaces, (ii) the background positive fixed charge  $N_{db}$ , and (iii) the inhomogeneous charge distribution associated with the ballistic motion of carriers injected at the contacts. The different resistance of the left and right branches due to the asymmetric profile of electron concentration along the horizontal branches, being higher near the negative electrode [Fig. 7(c)], is the consequence of the above effects. The surface charge lowers the electric potential when moving away from the contacts provoking the progressive depletion of the channel, thus leading to the typical minimum of potential and concentration in the middle of the structure, characteristic of space charge limited conditions [Fig. 7(d)]. When the TBJ is biased, the concentration shows an asymmetric shape (higher near the negative electrode due to the electron ballistic motion) leading to a shift of the potential minimum towards the negative electrode. As a consequence, the potential at the centre of the longitudinal channel is always negative (increasing with larger  $V$ ) and propagates to the bottom of the vertical branch, thus leading to the characteristic bell-shaped values of  $V_C$ . It is to be noted that we are using the constant surface charge model. In Fig. 7(a), we can also observe that the negative values of  $V_C$  reach a maximum for an intermediate value of  $\sigma$ , just when the width of the channel coincides with the lateral depletion induced by the surface charge (for  $\sigma/q = 0.25 \times 10^{12} \text{ cm}^{-2}$ ,  $W_d = 25 \text{ nm}$ , and the theoretical effective width of the channel  $W_{eff} = W - 2W_d$  becomes 0). In the next section we show how the self-consistent surface charge model (explained in section 2.4) introduce an extra asymmetry in the concentration leading to more negative values of  $V_C$ .

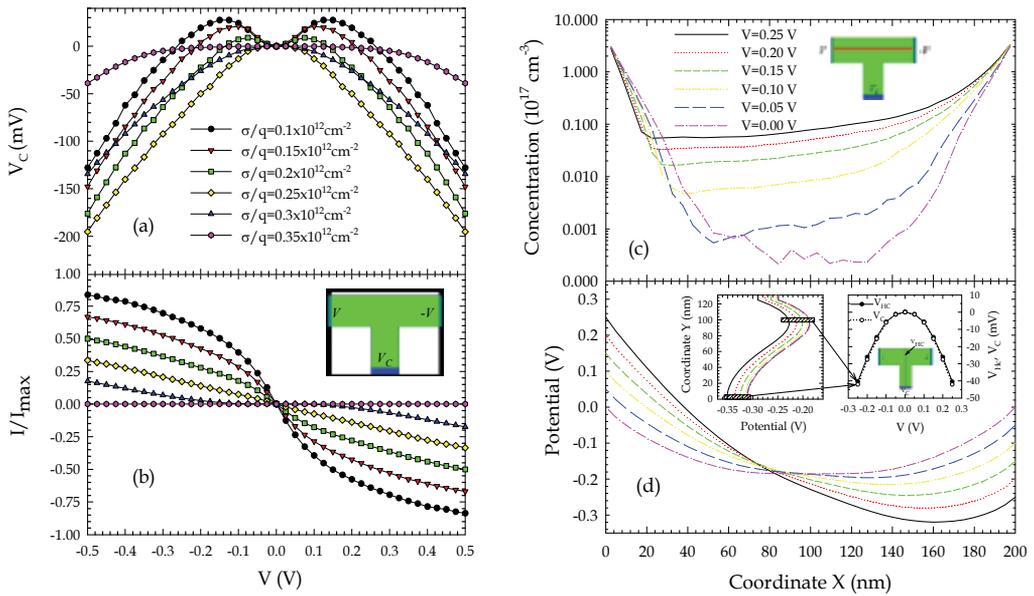


Fig. 7. (a) Electric potential at the bottom of the central branch of the TBJ and (b) normalized horizontal current when simulating with a constant surface charge model for different values. The inset shows the geometry of the TBJ with 50-nm-wide and 75-nm-long branches. (c) Horizontal electron concentration and (d) electric potential profiles of the TBJ for different bias conditions. The insets show the vertical potential profile in the middle of the central branch for several biasing, and the values of the potential at the bottom of this branch  $V_C$  and at the centre of the junction  $V_{\text{HC}}$  as a function of  $V$ .

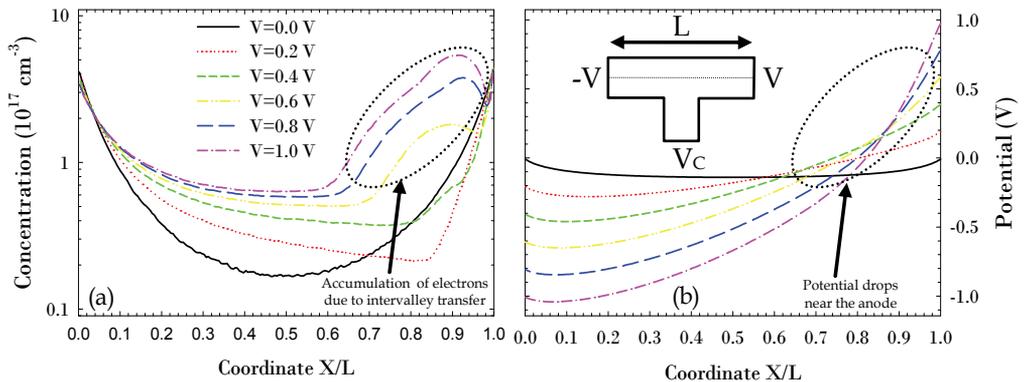


Fig. 8. (a) Profiles of electron concentration and (b) electric potential along the horizontal branch of the TBJ.

It is important to remark that, as sketched in Fig. 6(c), when increasing the bias ( $V > 0.25 \text{ V} \approx \Delta E_{\text{TL}}/2e$ ), even in the ballistic TBJs, the response of  $V_C$  becomes linear with  $V$  due to the appearance of  $\Gamma$ -L intervalley scattering mechanisms. That leads to the emergence of an accumulation domain near the positive electrode [Fig. 8(a)] completely screening the

variation of the potential drop between the central branch and the negative electrode [Fig. 8(b)]. This mechanism is also responsible for the saturation of the current [Fig. 7(b)] and the surprising negative values of  $V_C$  (since they are expected to be ohmic) measured in large TBJs [Mateos et al., 2004; Irie et al., 2008].

### 3.2 Self-consistent surface charge model. Stem width

As showed in experiments and contrary to the expectations, the central branch is not exactly a potential probe which measures the potential at the middle of the TBJ and therefore  $V_C$  depends on the stem width [Íñiguez-de-la-Torre et al., 2007]. Five TBJs with different widths of the vertical branch,  $W_{\text{VER}}=66, 78, 84, 94$  and  $108$  nm, have been fabricated and simulated in our MC tool with the self-consistent charge model (explained in section 2.4). Fig. 9(a) shows the MC values of  $V_C$  and the current  $I$  flowing through the horizontal branches in the TBJs when biased in push-pull fashion. The parabolic behaviour of  $V_C$  is strengthened when reducing the width of the vertical branch. This is in principle an unexpected result, since the vertical branch was believed to be only a measure (passive) element, in such manner that the value of  $V_C$  should be independent of its width  $W_{\text{VER}}$ . The inset of Fig. 9(a) shows how, as expected, the current is independent of  $W_{\text{VER}}$ , since the horizontal branch is identical for the different TBJs. The calculations performed with MC are consistent with the experimental results showing the same trend and a satisfactory quantitative agreement [Íñiguez-de-la-Torre et al., 2007].

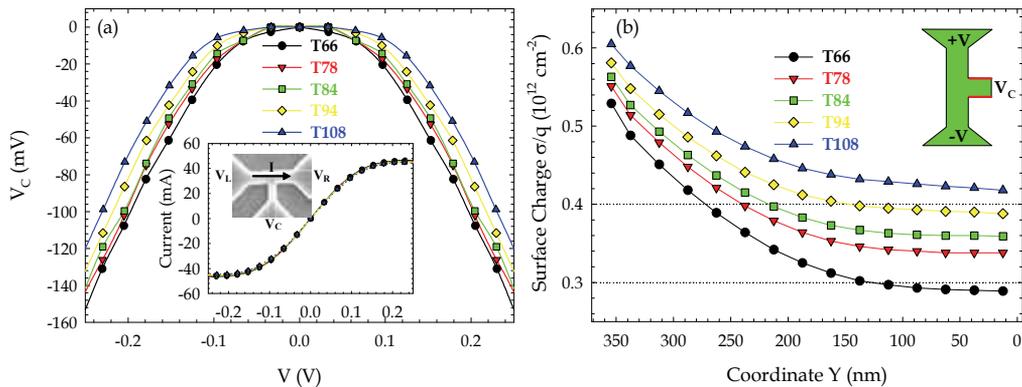


Fig. 9. (a) MC values of the bottom potential  $V_C$  and current (inset) in the TBJ junctions with 66, 78, 84, 94 and 108 nm wide vertical branches (denoted as T66, T78, T84, T94 and T108, respectively) as a function of the push-pull bias  $V$ . (b) Surface charges in the sidewalls of the vertical branch under equilibrium conditions ( $V=0$ ).

The proposed self-consistent surface charge model allows for the variation of the surface charge  $\sigma$  along the position in the interface in accordance with the surrounding free carrier concentration. This possibility of self-adaptation permits the surface charge in the narrowest junction (called T66) to reach values that approximately cause total depletion of the vertical branch,  $\sigma/q=0.3 \times 10^{12} \text{ cm}^{-2}$ . In contrast, in the widest junction (called T108) the surface charge is limited by a value close to that obtained in the experimental measurements,  $\sigma/q=(0.4 \pm 0.1) \times 10^{12} \text{ cm}^{-2}$  [Fig. 9(b)]. Furthermore, the surface charges take a value practically constant near the bottom of the vertical branch, which indicates that the results will not change if this branch is made longer.

In previous works, within the constant charge model, the vertical branch was considered as a voltage probe, providing at its bottom ( $V_C$ ) the variations of  $V_{HC}$  (potential at the centre of the horizontal branch) [inset 7(d)]. However, within the self-consistent model, the surface charge at the sidewalls of the vertical branch and the carrier penetration inside it change with the bias (like in Y-Junctions), which provokes that  $V_C$  is no longer a faithful reflection of the  $V_{HC}$  variations and it is necessary to consider the electric potential difference  $\Delta V_V$  between  $V_C$  and  $V_{HC}$ , which is not equal for all the biasing. Therefore, the values of  $V_C$  can be considered as the result of two combined effects: a horizontal one (given by  $V_{HC}$ ) and a vertical one (given by  $\Delta V_V$ ).

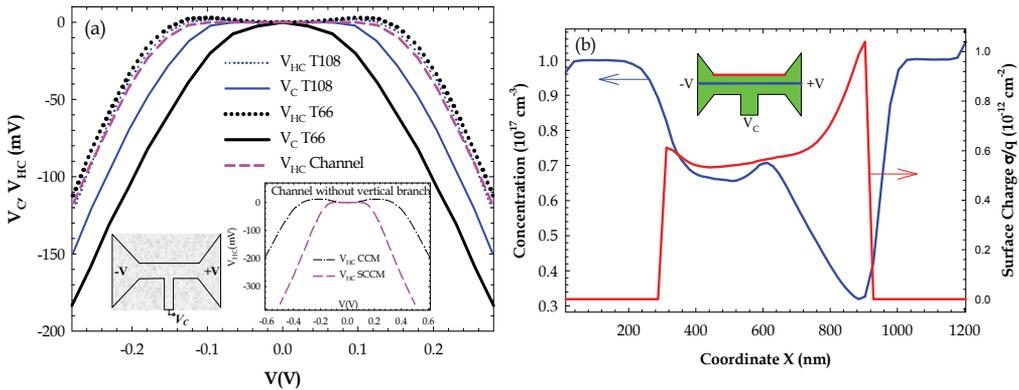


Fig. 10. (a)  $V_{HC}$  and  $V_C$  as a function of the applied voltage for the TBJs with  $W_{VER}=66$  and  $108 \text{ nm}$ .  $V_{HC}$  for a channel without the vertical branch is also plotted for comparison. The inset shows  $V_{HC}$  calculated in the channel without the vertical branch with the constant charge model ( $\sigma/q=0.4 \times 10^{12} \text{ cm}^{-2}$ ) and self-consistent charge model. (b) Profiles of carrier concentration along the centre of the horizontal branch and surface charge along the top boundary of the T66 for  $V=V_R=-V_L=0.25 \text{ V}$ .

We analyze the horizontal effect, that is, the values of  $V_{HC}$ . In Fig. 10(a) we plot the values obtained with the MC simulations for  $V_{HC}$  in T66 and T108, together with those calculated in a channel without vertical branch. It can be observed that, as expected, the values of  $V_{HC}$  practically coincide in the three structures (transport takes place in horizontal direction and the width and length of the horizontal branch are practically the same). The inset of Fig. 10(a) also compares the values of  $V_{HC}$  in a simple channel (without vertical branch) obtained with the constant and self-consistent charge models. The self-consistent model leads to a considerable enhancement of the negative values of  $V_{HC}$ , which is the signature of an enhanced electron charge asymmetry in the horizontal direction, especially for high biasing, as can be shown in Fig. 10(b) for  $V=0.25 \text{ V}$ . A strong depletion of carriers takes place at the anode side of the horizontal branch. This region becomes highly resistive, so that most of the applied potential drops here and leads to the high negative values of  $V_{HC}$ . The origin of the increase of the surface charge with the applied voltage near the anode lies in the fact that, due to the ballistic motion of electrons, their longitudinal energy increases significantly as they approach the right contact. Scattering mechanisms, even if there are very few, produce some energy redistribution, and thus make the transversal energy component also increase. In this way electrons are able to approach the boundaries of the TBJ (in spite of the repulsive effect of the surface charge) and contribute to raise the value of  $\sigma$ .

### 3.3 Influence of the horizontal branch

Our self-consistent surface charge model can also be applied to explain the behaviour when the length  $L_{\text{HOR}}$  and the width  $W_{\text{HOR}}$  of the horizontal branches are modified [Fig. 11]. When the length of the horizontal branches  $L_{\text{HOR}}$  is changed we found that the values of  $V_{\text{HC}}$  are very similar for all the structures. This is due to the analogous horizontal concentration profiles found for the different lengths. As in the previous TBJs, it is the presence of the vertical branch and the associated surface charges which leads to different values of  $V_{\text{C}}$  in each of the structures. Like in the case of the experimental results [Irie et al., 2008] we found that the down-bending behaviour of  $V_{\text{C}}$  is stronger for shorter junctions. This result is expected because of the more ballistic character of transport in shorter structures, but our results indicate that surface charges and the presence of the vertical branch also play a role.

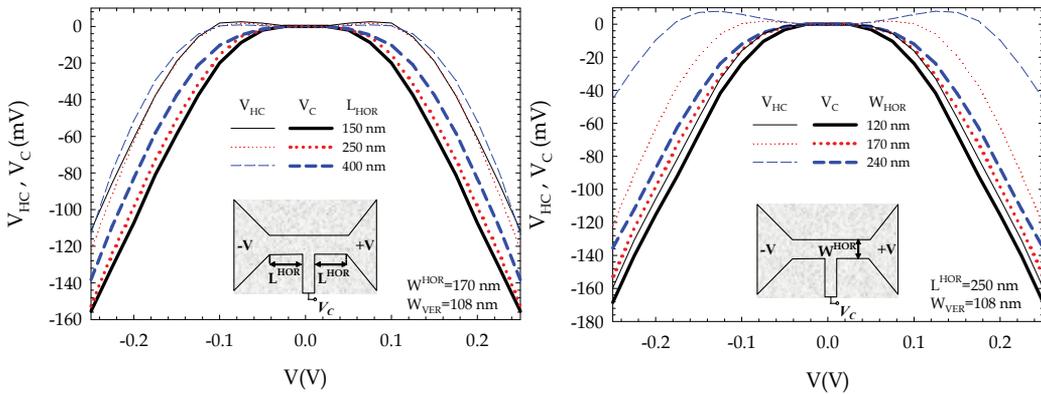


Fig. 11. (a)  $V_{\text{HC}}-V$  and  $V_{\text{C}}-V$  for TBJs with  $L_{\text{HOR}}=150, 250,$  and  $400$  nm. (b)  $V_{\text{HC}}-V$  and  $V_{\text{C}}-V$  for TBJs with  $W_{\text{HOR}}=120, 170,$  and  $240$  nm.

Concerning the width of the horizontal branch  $W_{\text{HOR}}$ , the values of  $V_{\text{C}}$  are higher (more negative) as the width is decreased, in accordance with the trend found in our experiments [Íñiguez-de-la-Torre et al., 2009a]. However, since the length is identical, the origin must be related not only to ballistic transport. Results can be interpreted in terms of a new factor: the strength of surface charge effects. Remarkably, and in contrast with the behaviour found when modifying  $W_{\text{VER}}$  and  $L_{\text{HOR}}$ , in this case the values of the potential at the centre of the junction  $V_{\text{HC}}$  exhibit a strong dependence on  $W_{\text{HOR}}$ . The narrower the horizontal branch, the lower the free carrier concentration due to the stronger depletion induced by the enhanced surface charge. For this reason the horizontal potential profile is different in each junction and also the value of  $V_{\text{HC}}$ . The dependence of  $V_{\text{HC}}$  on  $W_{\text{HOR}}$  is smoothed in the bottom potential  $V_{\text{C}}$  by surface charge adaptation in the vertical branch as explained in the previous section. Therefore, it is the role played by surface charge and the associated depletion what leads to the observed variations between structures with different  $W_{\text{HOR}}$ . The surface charge influence on  $V_{\text{C}}$  is much stronger than that of ballistic transport in the case of small  $W_{\text{HOR}}$ , but decays sharply for wide TBJs in agreement with the experimental findings [Íñiguez-de-la-Torre et al., 2009a]. Finally we want to remark that in the wider TBJs, for the lowest applied voltages,  $V_{\text{HC}}$  gets slightly positive values before becoming negative. MC microscopic results indicate that under such bias conditions the branch contacted to the cathode is more resistive than the anode one providing these positive values. It is due to a

velocity overshoot effect (more pronounced near the cathode) taking place for weak enough surface charge effects.

Overall, we can conclude that the parabolic response (associated with ballistic transport) sharply decreases for larger  $L_{HOR}$ , and increases in efficiency for narrower channels ( $W_{VER}$  and  $W_{HOR}$ ). Experiments also shown an unexpected significant degree of ballistic transport even in TBJs with  $L_{HOR}=2 \mu\text{m} \gg l_m$  [Irie et al., 2008] and  $W_{HOR}=2 \mu\text{m}$  [Íñiguez-de-la-Torre et al., 2009a]. This robustness of the TBJs nonlinear response is a unique advantage in terms of realizing a room-temperature integrated circuitry using TBJs over a wide range of channel sizes.

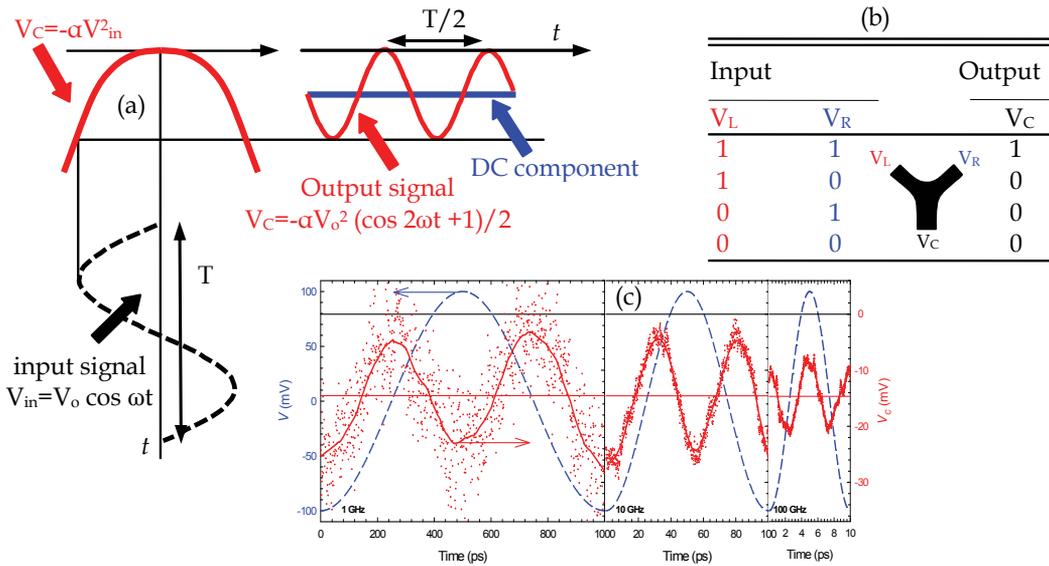


Fig. 12. (a) Illustration of detection and frequency doubling of RF signals in TBJs, (b) logic truth table for a symmetric TBJ device, operated as an AND gate, when defining positive voltage as the binary value of 1 and (c) MC time domain simulation of a TBJ for different frequencies of a push-pull input signal of amplitude 100 mV: 1, 10 and 100 GHz.

### 3.4 Rectification and doubling in TBJs

The nonlinear response of TBJs can be exploited to perform several analog and digital functions as illustrated in Fig. 12. An example is high-frequency signal rectification [Bednarz et al., 2005] and second-harmonic generation [Lewén et al., 2002]. Frequency mixing, doubling and phase detection have also been demonstrated [Shorubalko et al., 2002; Sun et al., 2007; Gardès et al., 2008]. Another example of application is as logic gate for digital electronics [Rahman et al., 2009]. It is clear, for example [Fig. 12 (b)], that if we use left and right branches as inputs, the central branch output will perform the logic AND operation ( $V_C$  has high voltage only when both  $V_L$  and  $V_R$  are high and low voltage in other cases) [Xu, 2002]. It was reported that with more complicated structures of several TBJs can work as NAND [Reitzenstein et al., 2002, Xu et al., 2004], NOR [Müller et al., 2007], half adder [Worschech et al., 2003; Reitzenstein et al., 2004], full adder [Lau et al., 2006] or SR latch [Sun et al., 2008], etc.

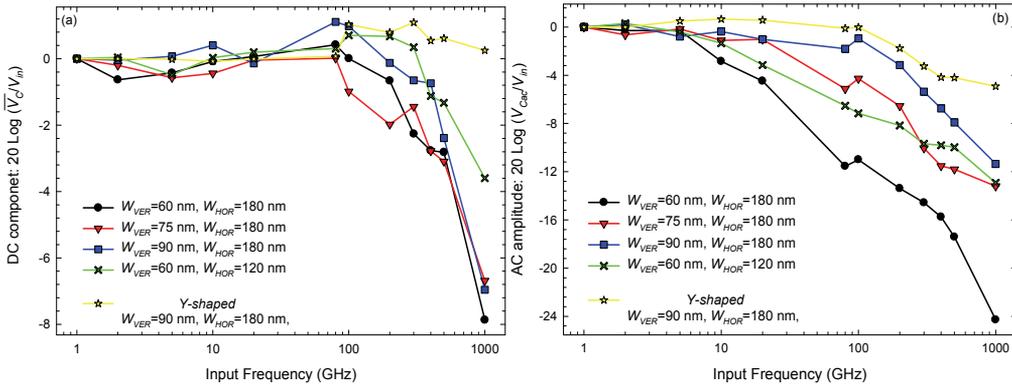


Fig. 13. Frequency dependence of (a) amplitude,  $V_{Cac}$  and (b) average DC value,  $\overline{V_C}$ , of the response of  $V_C$  (normalized to the input amplitude and in dB) to signals of amplitude 100 mV applied in push-pull to the inputs in T-shaped junctions with different  $W_{VER}$  and  $W_{HOR}$  and in a Y-shaped junction. For a better comparison, the low-frequency value has been subtracted to each curve.

In this section we study the effect of the geometry ( $W_{HOR}$ ,  $W_{VER}$ ) and shape on the performance of three-branch junctions operating as detectors and frequency doublers [Fig. 12(a)]. To this end, we apply 100 mV sinusoidal signals in push-pull fashion. In Fig. 12(c) the time domain evolution of the stem output voltage  $V_C$  is represented when the frequency of the input signal is 1, 10 and 100 GHz. It can be observed that the device has an excellent performance as frequency doubler at least up to 100 GHz. Fig. 13 shows the values for the amplitude,  $V_{Cac}$  and the average DC value  $\overline{V_C}$  of the response of  $V_C$ , as a function of frequency for the different simulated junctions (for the sake of clarity the low-frequency value has been subtracted). As a general feature, the cut-off for  $V_{Cac}$  appears at much lower frequencies than for the mean value,  $\overline{V_C}$ , taking place around 1 THz [Fig. 13(a)]. For the T-shaped junctions as the value of  $\overline{V_C}$  is mainly related to the electron horizontal transport, its cut-off is hardly influenced by the width of the vertical branch. On the other hand,  $V_{Cac}$  is controlled by the penetration of carriers into the stem, so that its width ( $W_{VER}$ ) clearly changes its cut-off frequency (higher frequencies for wider stems, in which carriers enter more easily). Nevertheless, wider stems provide less negative values of  $V_C$  at low-frequency. As a consequence, for an optimized response  $W_{VER}$  must be chosen depending on the required type of operation and frequency. On the other hand, when reducing  $W_{HOR}$ , a slight increase of the cut-off frequency is observed in both quantities; however, the matching to the typical 50  $\Omega$  lines would be worse due to the higher impedance of the TBJ. Concerning the shape, the Y geometry much improves the global performance of the device as a result of an enhanced vertical electric field and a stronger injection of carriers into the stem (having a more pronounced influence on the cut-off of  $V_{Cac}$ , associated with frequency doubling applications). Moreover, better performances are expected if the angle between left and right branches of the junction is further decreased.

It is to be noted that in these high-frequency MC simulations the profile of the surface charge is frozen to that previously calculated under equilibrium conditions, since the typical capture/emission times of the surface states are about 1  $\mu$ s, several orders of magnitude longer than the maximum period used for exciting the junctions (1 ns). In fact, a low

frequency plateau should be observed in experiments with a cut-off corresponding to the inverse characteristic lifetime of the surface charge traps.

#### 4. Self-switching diodes

The second nanometer-scale nonlinear device that we study in detail in this chapter was proposed in 2003 by A. M. Song and named Self-Switching Diode. The key point in the fabrication of the device is the etching of the two L-shaped insulating grooves defining a narrow semiconductor channel [Fig. 1(b)]. An applied voltage  $V$  not only changes the potential profile along the channel direction, but also either widens or narrows the effective channel depending on the sign of  $V$ . This results in a diode-like characteristic, but without the use of any doping junction or barrier structure.

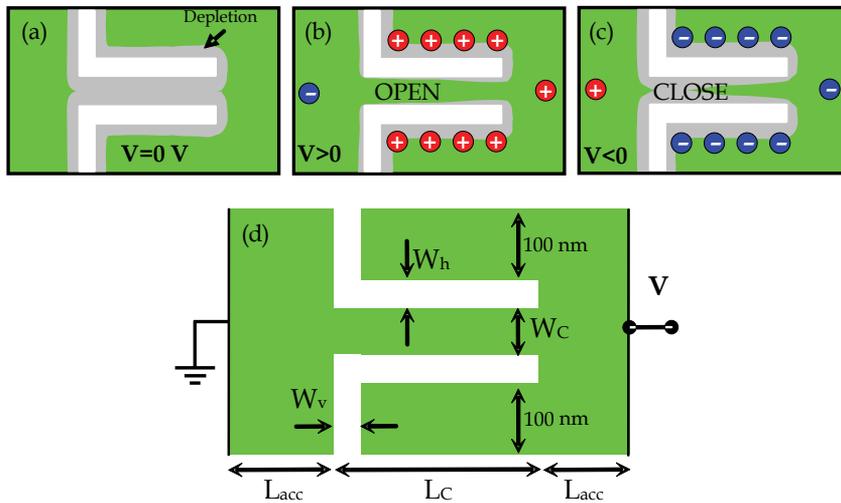


Fig. 14. (a) Depletion region formed close to the etched boundaries in equilibrium. Depending on the sign of the applied voltage the effective channel width will (b) increase or (c) reduce. (d) SSD geometry for the MC simulations.

The two-terminal structure allows SSD-based circuits to be realized by a simple single step lithographic process, so that its size can be easily reduced to the nanometer-range. Thus, by using fast III-V materials, the high frequency performance of SSDs can be dramatically boosted thanks to a much shorter transit time, due not only to a smaller channel length but also to an enhanced electron velocity associated to ballistic transport. And last but not least, the planar geometry of the SSDs allows placing the two contacts with long separation, so that parasitic crosstalk capacitances can be drastically reduced. These facts, together with the intrinsically high electron velocity channels, should permit the fabrication of SSDs working in the THz range.

##### 4.1 Overview, working principle and I-V curves

The voltage applied to the anode (right contact) of the SSD propagates to the vicinity of the channel, while in the cathode region (at the left of the trenches) the potential is always essentially zero. In equilibrium, the channel is closed due to the depletion induced by the

surface charges located at the lateral walls [Fig. 14(a)], which lead to the appearance of a longitudinal potential barrier. When  $V > 0$ , the positive voltage reaches the lateral regions of the SSD channel, so that the potential barrier is lowered (or even removed), thus allowing the electron flow (the channel is open) [Fig. 14(b)]. On the contrary, when  $V < 0$ , the potential profile in the right part of the device is almost unchanged with respect to the equilibrium situation (it is just shifted to lower values), the channel thus remaining closed [Fig. 14(c)].

We analyze I-V curves, noise spectra and rectification when some parameters of the diode geometry are modified (keeping the others constant). The reference SSD for all the simulations will be the one with:  $W_C = 50$  nm,  $L_C = 250$  nm,  $W_h = W_v = 5$  nm and  $L_{acc} = 175$  nm. In general terms the forward current shows an exponential dependence on the applied voltage for low values of  $V$  (as long as the barrier is present), and then becomes linear (resistive behaviour), with a tendency to saturation at the highest applied voltages due to hot-carrier effects [insets Fig. 16].

For devices with smaller channel width,  $W_C$ , inset Fig. 16(a), the turn-on voltage is larger due to a larger barrier at equilibrium, which needs a higher applied voltage to disappear. The length of the channel,  $L_C$ , is also found to largely influence the device behaviour. As observed in the inset of Fig. 16(b), short-channel effects appear when the aspect ratio of the channel ( $L_C/W_C$ ) decreases. In such a case, under reverse bias, the potential of the lateral regions is not able to deplete the channel, so that the barrier preventing the current flow disappears and an inverse leakage current flows (as observed for  $L_C = 100$  nm). The very thin trenches of the devices not only prevent the presence of inverse leakage current for very short channels (it only appears for  $L_C = 100$  nm), but also the forward current is much improved, inset Fig. 16(c). As observed in the inset of Fig. 16(c) and Fig. 16(d), the turn-on voltage is influenced by the width of the horizontal (and not by the vertical) trenches, decreasing for smaller  $W_h$ . This is due to the stronger transverse electric field present for smaller  $W_h$ , which enables a more efficient control of the opening and closing of the nanochannel when biasing the anode.

Therefore, the operation principle of this device is similar to that of an enhanced mode field effect transistor (pinched off at equilibrium) in which lateral gates (in this case short circuited to the drain) control the current flow through the channel. From the point of view of applications, the non-linear response of the diodes and the ultra fast ballistic transport opens the possibility of fabricating circuits for rectification, detection or even harmonic generation at very high frequencies.

#### 4.2 Noise spectra

By downscaling the device dimensions, the detection properties of an array of SSDs has been proved up to 110 GHz at RT [Balocco et al., 2005] and up to 2.5 THz at 10 K [Balocco et al., 2008]. At so high frequencies, the intrinsic noise generated by the diodes becomes a performance limitation, and must be carefully analyzed in order to reduce its level as much as possible. Initial studies [Íñiguez-de-la-Torre et al., 2008] of the spectral density of current fluctuations at low-frequency  $S_I(0)$  (in the plateau beyond the  $1/f$  range) compared to the  $2qI$  value (with  $q$  the electron charge) have evidence the presence of full shot noise not only under inverse but also under direct bias below a certain threshold potential, above which the level of noise is lower than  $2qI$  [Fig. 15(a)]. Full shot noise (both under forward and reverse bias) appears when the transport is barrier controlled and the current is provided by uncorrelated carriers surpassing the barrier. The noise temperature is close to half the value

of the lattice temperature associated with an ideal exponential dependence of the forward current, which usually goes along with the previously commented full shot-noise behaviour. At high forward bias the barrier is lowered or even disappears, the channel resistance decreases, and the diffusive accesses to the channel become more important and the noise temperature increases significantly over the lattice temperature due to a strong electron heating.

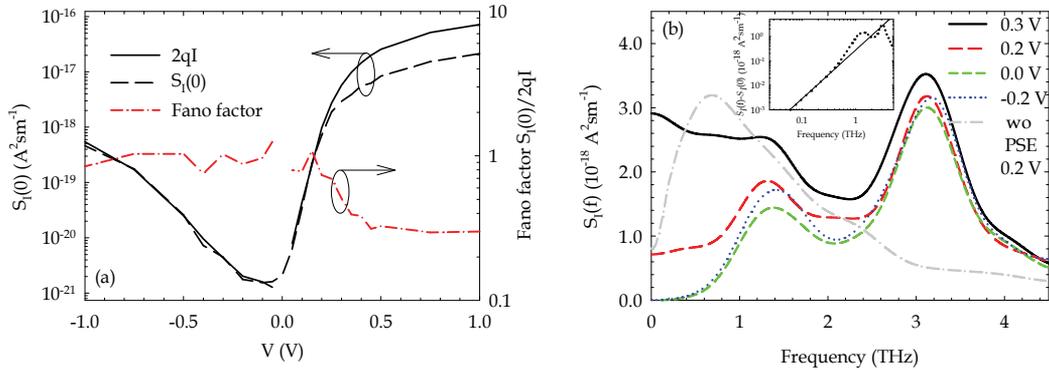


Fig. 15. (a) MC values of  $S_I(0)$ , compared to  $2qI$ , as a function of the applied voltage (left axis). The Fano factor  $S_I(0)/2qI$  is also plotted (right axis). The inset shows the I-V curve of the diode in linear scale. (b)  $S_I(f)$  for several bias conditions in the reference SSD. The case in which the PS is switched off is also shown for  $V=0.2$  V. The inset illustrates the  $f^2$  dependence of the noise spectrum for  $V=0.0$  V. Reference SSD.

For high frequency the current noise spectra  $S_I(f)$  for the reference SSD is shown in the Figure 15(b). Two main peaks are observed in the spectra. 3D plasma oscillations are at the origin of the one appearing at the highest frequencies (above 3 THz). When the PS is switched off plasma oscillations are not present and this peak disappears. The other peak, between 1-2 THz, is attributed to returning-carrier effects taking place in the space-charge regions originated by the surface charge at both sides of the vertical trenches. It exhibits the characteristic  $f^2$  behaviour [inset of Figure 15(b)] already found in other devices like Schottky-barrier diodes, revealing a capacitive coupling of the returning carrier fluctuations to the noise at the terminals. But more interesting, and in contrast with the other peak, the frequency of this one depends on the geometry of the SSD as we can see in Figure 16.

As observed, the level of noise at high frequency is higher the larger is the impedance of the accesses as compared to that of the channel. This explains, for example, why  $S_I(f)$  is higher when decreasing  $L_C$  or increasing  $W_C$  or  $L_{acc}$ , while it remains with similar amplitude when changing  $W_h$  or  $W_v$ . The increase of  $S_I(f)$  originated by the peak at lower frequency could limit the frequency range of potential applications. Thus, a first possibility to reduce the noise related to the returning-carriers peak is to decrease the resistance of the accesses relative to that of the channel. The best choice to this end is shortening  $L_{acc}$  [see Fig. 16(e)], which is always desirable to reduce parasitic resistances but may increase the parasitic capacitance between electrodes. Decreasing  $W_C$ , Fig. 16(a), or increasing  $L_C$ ; Fig. 16(b), would enhance the channel resistance and reduce the current level, both undesirable effects. Moreover, a longer channel leads to lower cut-off frequency. A second possibility is to try to move the peak to higher frequencies, thus reducing the amplitude of the noise in the range

of interest (around 1 THz). As observed in Fig. 16, by modifying  $W_C$ ,  $L_C$  or  $L_{acc}$ , the frequency of the maximum hardly changes. In contrast, an increase in the width of the vertical trenches  $W_v$ , shifts the peak to higher frequencies, Fig. 16(d).

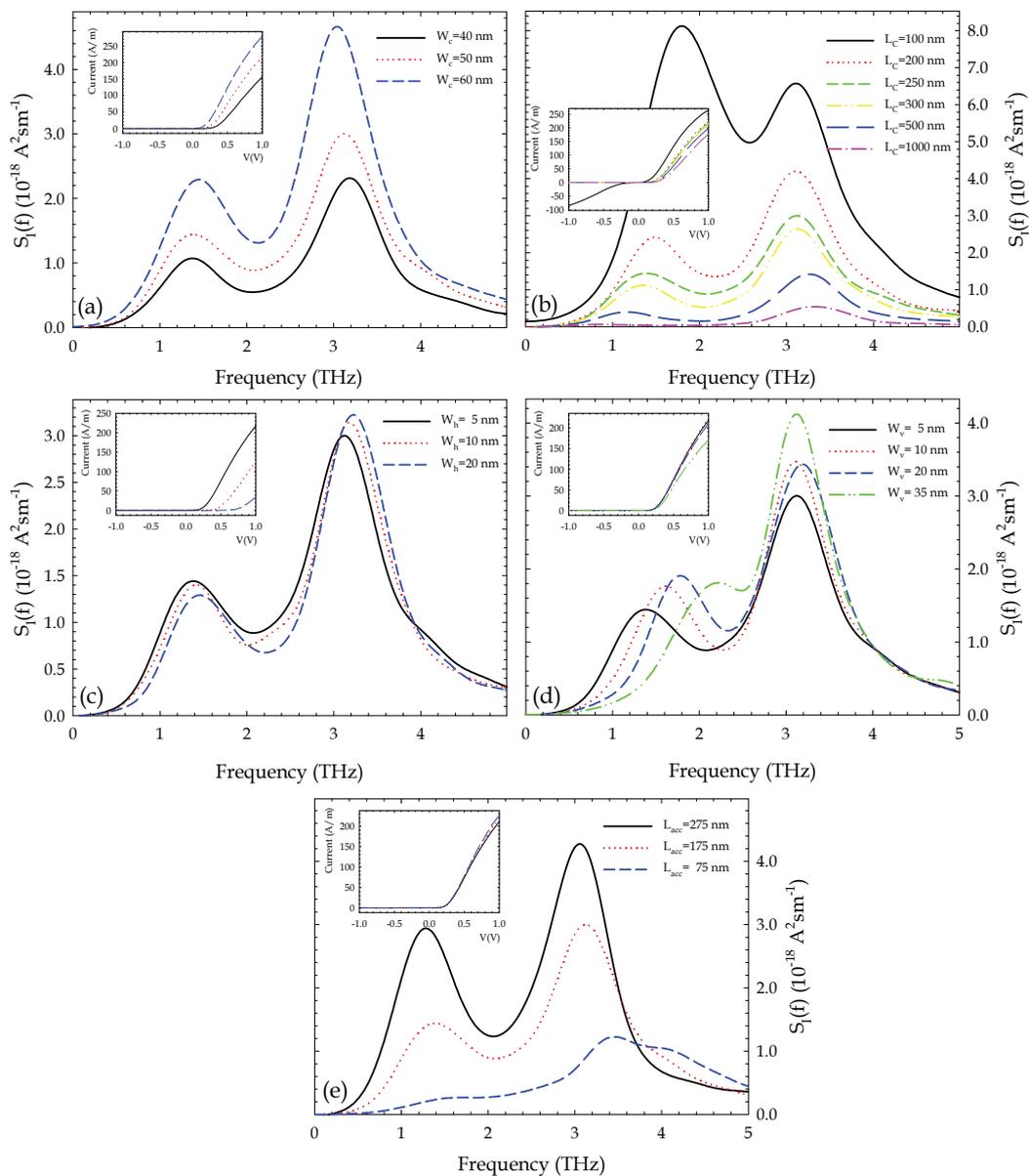


Fig. 16. Current-noise spectra at equilibrium when some parameters of the topology of the diode are modified: (a)  $W_C=40, 50$  and  $60 \text{ nm}$ , (b)  $L_C=100, 200, 250, 300, 500$  and  $1000 \text{ nm}$ , (c)  $W_h=5, 10$  and  $20 \text{ nm}$ , (d)  $W_v=5, 10, 20$  and  $50 \text{ nm}$  and (e) length of the accesses  $L_{acc}=275, 175$  and  $75 \text{ nm}$ . The insets show the corresponding I-V characteristics.

### 4.3 Rectification to AC signals

We have analyzed the dynamic behaviour of SSDs in terms of their AC to DC rectification (RF detection) as the main application of the device [Íñiguez-de-la-Torre et al., 2009b]. Harmonic voltage signals  $V=V_0\sin(2\pi ft)$  of increasing frequency  $f$  are applied between the contacts and the mean value of the output current is evaluated for the same “set” of SSDs than in the previous section [Fig. 17]. After a flat region at the lower frequencies, the rectified current exhibits a pronounced peak,  $f_p$ , just before the decay in the response. As expected,  $f_p$  depends on the channel length  $L_C$ , Fig. 17(a), being lower for longer channels. The SSD with  $L_C=100$  nm is correctly responding up to frequencies over 2.0 THz, thus making possible the operation of these devices as, for example, power detectors of THz waves. We also observe that  $f_p$  is highly sensitive to the properties of the vertical trench width  $W_v$  [Fig. 17(c)] but insensitive to that the horizontal trenches,  $W_h$  [Fig. 17(b)].

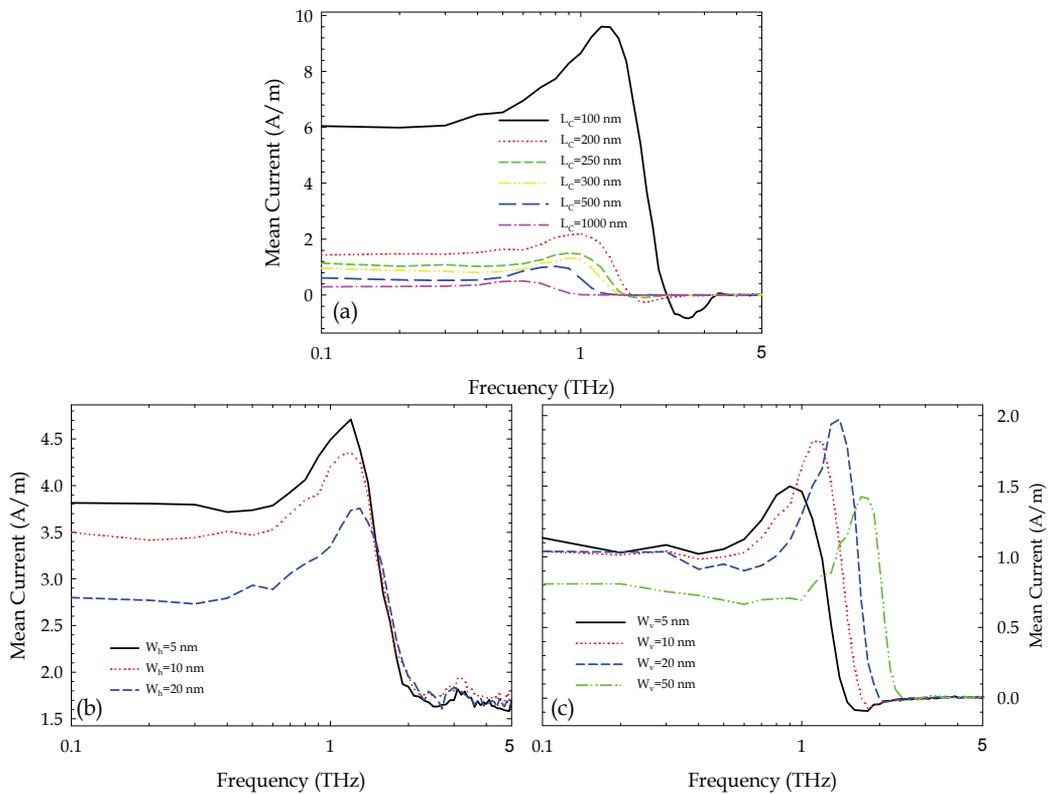


Fig. 17. Mean response current to a periodic input voltage (with amplitude of 0.25 V) applied to SSDs: (a)  $L_C=100, 200, 250, 300, 500$  and  $1000$  nm, (b)  $W_h=5, 10$  and  $20$  nm, (c)  $W_v=5, 10, 20$  and  $50$  nm.

Significantly, the behaviour of the low frequency peak in the noise spectra and its relative amplitude [Fig. 16] is in principle surprisingly similar to the  $f_p$  corresponding to the dynamic response of the rectified current [Fig. 17]. This indicates that we are observing the same microscopic phenomenon reflected in different macroscopic quantities. Let us explain this behaviour. The origin of the low frequency peak and  $f_p$  is related to the dynamics of the

reflected carriers (returning carriers) in the region close to both sidewalls of the vertical trenches, which capacitive couple to the current at the terminals. This coupling is modulated by two capacitors, one associated to the depletion regions at both sides of the trenches, and the other due to the vertical trenches themselves,  $C_v = \epsilon_v / W_v$  ( $\epsilon_v$  is the permittivity of the vertical trench). A more detail analysis of the effect of the insulator material filling the trenches can be found in [Íñiguez-de-la-Torre et al., 2009b]. That means that a typical noise mechanism, which is a collective charge fluctuation, is coupled due to the geometry provoking a resonant peak  $f_p$  in the AC to DC rectification. However the rectified current is phase shifted by the horizontal capacitor, while the noise is only influenced by the vertical one (with no need of the presence of a conducting channel), thus explaining why the frequency of the peak does not coincide exactly in both quantities. In addition the spectral density at equilibrium is proportional to the small signal admittance of the device, whereas the DC response current is caused by the diode rectification and corresponds to large signal conditions.

**4.4 InAs and InSb SSDs**

The quite useful tuneable-by-geometry detection in the terahertz range observed in InGaAs SSD exhibit however a low amplitude and quality factor. In this section we will show how the low effective mass of InAs and InSb in relation to InGaAs enhances ballistic transport inside the diode, thus improving the detection sensitivity. A clear enhancement in the resonance and shifting  $f_p$  to higher frequencies is observed for these two narrow band gap materials [Fig. 18]. The resonance in the rectified DC current exhibits a remarkable quality factor, much higher than in InGaAs, with amplitude more than a hundred times the low frequency value in the case of InSb SSDs and also at higher frequency tuning range.

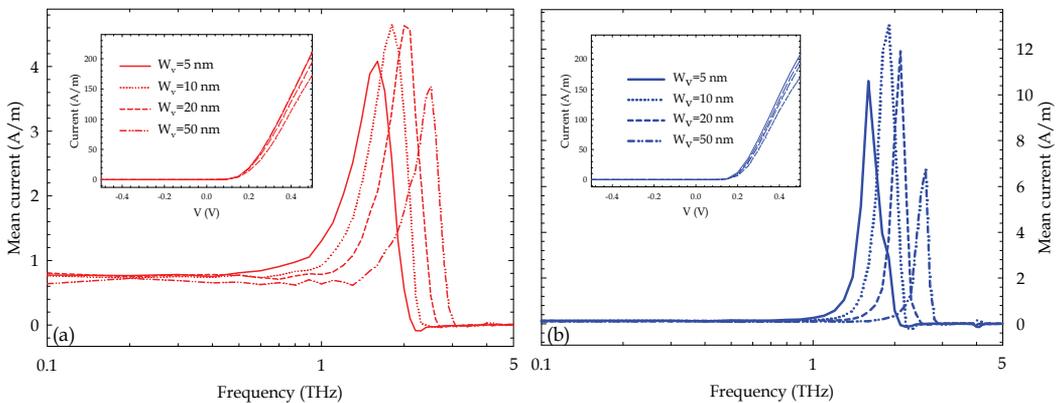


Fig. 18. Mean current response to a periodic input voltage (with amplitude of 0.15 V) applied to InAs-based SSD and (b) InSb-based SSD. The insets show the corresponding I-V curves.

To corroborate the clear link between the peak in the rectified DC current and the similar one present in the noise spectral density, in Fig. 19(a) we plot the noise spectra at equilibrium for SSDs based on three high mobility semiconductors: InGaAs, InAs and InSb (for  $W_v = 20$  nm). In Fig. 19(b), the frequency form noise peaks together with  $f_p$  of Fig. 17(c), 18(a) and 18(b) are plotted for each material. The three frequencies exhibit a similar relative

variation between the different materials, mainly affected by their effective masses and dielectric constants (0.014 and 17.65, respectively, for InSb, 0.023 and 15.15 for InAs and 0.042 and 13.88 for  $\text{In}_{0.47}\text{Ga}_{0.53}\text{As}$  [Rodilla et al., 2009]) thus revealing a common origin linked with plasma effects. Remarkably, the peak in the rectified DC current has the same tendency, certifying our previous conjecture that a noise mechanism is coupled via the particular geometry of the SSD to the DC to AC response and enhances the performance of the device.

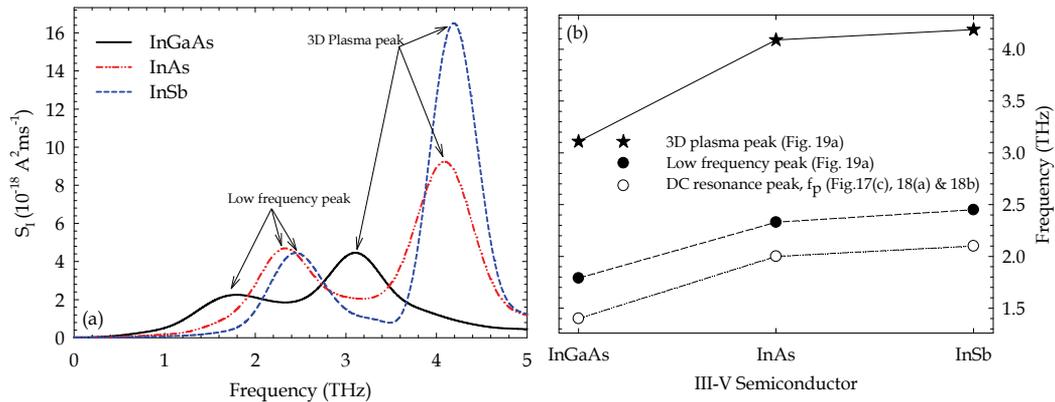


Fig. 19. (a) Current noise spectra at equilibrium for InGaAs, InAs and InSb diodes ( $W_V=20$  nm). (b) Comparison of the frequency peaks in the noise spectra with those of Fig. 17(c), 18(a) and 18(b).

## 5. Conclusion

Nowadays fabrication of narrow channels with a length of few nanometers is possible. Electrons move without suffering losses in their momentum and energy, leading to several new transport phenomena like ballistic transport. Our MC model treats the electrons in a billiard ball-like manner (classical), but it locally takes into account the effect of degeneracy by using the rejection technique (the scattering mechanisms are rejected when the final energy state is likely to be occupied). It is to be noted that as our study is focused on room temperature operation, quantum effects such as energy quantization can be initially neglected.

Monte Carlo simulations have significant advantages over other methods: there are no adjustable parameters (only material related microscopic parameters), it gives precise information about internal quantities (including scattering) and it is able to describe non-local non-static processes like ballistic transport and collective phenomena like plasma oscillations. In summary, it is able to provide static, dynamic and noise results with the only drawback of CPU intensive and slow simulations.

Monte Carlo simulations have shown that ballistic nanodevices like TBJs and SSDs have unique and striking high-frequency operation capability. However, there are significant problems to be solved. The reduction in length (so that the transport is almost purely ballistic) and width of the branches (enhancing the space charge effects), to optimize the performance of applications, greatly increases the impedance of the devices (to the range of  $k\Omega$ ). As a consequence even very small parasitic capacitances (of the order of fF) prevent the

extrinsic cut-off frequencies of the devices to reach the THz range. Therefore, design guidelines for the reduction of the device parasitics, together with other ways of improvement as the use of more than one device in parallel must be used.

From a technological point of view the compatibility of these ballistic devices with the well-established high electron mobility transistors (HEMTs) circuitry should bring versatility for many practical applications, i.e. by allowing the processing of the signals generated by the ballistic devices. However, the final objective to potentially replace the conventional CMOS design flow and enable high performance analog and digital circuit design using this *family* of ballistic nanodevices, will require much more efforts directed towards the development of a logic compatible device, smaller interconnections, lower routing complexity, higher fabrication throughput, higher reliability, etc.

## 6. References

- Balocco C., Song A. M., Aberg M., Forchel A., González T., Mateos J., Maximov I., Missous M., Rezaazadeh A. A., Saijets J., Samuelson L., Wallin D., Williams K., Worschech L. & Xu H. Q. (2005). Microwave detection at 110 GHz by nanowires with broken symmetry, *Nano Letters* 5, 1423.
- Balocco C., Halsall M., Vinh N. Q. & Song A. M., (2008). THz operation of asymmetric-nanochannel devices, *J. Phys.: Condens. Matter* 20, 384203.
- Bednarz L., Rashmi, Hackens B., Farhi G., Bayot V. & Huynen I. (2005). Broad-band frequency characterization of double Y-branch nanojunction operating as room-temperature RF to DC rectifier, *IEEE Trans. Nanotechnol.* 4, 576.
- Bednarz L., Rashmi, Simon P., Huynen I., González T. & Mateos J. (2006). Negative differential transconductance and nonreciprocal effects in Y-branch nanojunction. High-frequency behavior, *IEEE Trans. Nanotech.* 5, 750.
- Büttiker M. (1986). 4 terminal phase coherent conductance, *Phys. Rev. Lett.* 57, 1761 (1986).
- Cappy A., Bayot V., Bednarz L., Bollaert S., Boutry H., Gonzalez T., Hackens B., Huynen I., Gallo J. S., Mateos J., Pardo D., Rashmi, Roelens Y., Vasallo B. G. & Wallart X. (2005). Ballistic Nanodevices for Terahertz Data Processing, Nanotera EU IST-2001-32517 project 3rd year review report. Tech. Rep., Nanotera Consortium, February 2005, available online at: [http://www.phantomsnet.net/files/EUprojects/NANO\\_TERA\\_FR.pdf](http://www.phantomsnet.net/files/EUprojects/NANO_TERA_FR.pdf).
- Diduck Q., Irie H. & Margala M. (2009). A room temperature ballistic deflection transistor for high performance applications, *Int. J. of High Speed Electron. Syst.* 19, 23-31.
- Galloo J. S., Pichonat E., Roelens Y., Bollaert S., Wallart X., Cappy A., Mateos J. & González T. (2004). Transition from ballistic to ohmic transport in T-branch junctions at room temperature in GaInAs/AlInAs heterostructures, *Proc. of the 2004 International Conference on Indium Phosphide and Related Materials IPRM*, IEEE Catalog 04CH37589, 378.
- Gardès C., Roelens Y., Bollaert S., Galloo J. S., Wallart X., Curutchet A., Gaquiere C., Mateos J., González T., Vasallo B. G., Bednarz L. & Huynen I. (2008). Ballistic nanodevices for high frequency applications, *Int. J. Nanotechnology.* 5, 796.
- González T. & Pardo D. (1996). Physical models of ohmic contact for MC device simulation, *Solid-State Electron.* 39, 555.

- González T., Mateos J., Pardo D., Varani L. & Reggiani L. (1999). Injection statistics simulator for dynamic analysis of noise in mesoscopic devices, *Semicond. Sci. Technol.* 14, L37.
- González T., Vasallo B. G., Pardo D. & Mateos J. (2004). Room temperature nonlinear transport in ballistic nanodevices, *Semicond. Sci. Technol.* 19, S125.
- Hieke K. & Ulfward M. (2000). Nonlinear operation of the Y-branch switch: Ballistic switching mode at room temperature, *Phys. Rev. B* 62, 16727.
- Íñiguez-de-la-Torre I., Mateos J., González T., Pardo D., Galloo J. S., Bollaert S., Roelens Y. & Cappy A. (2007). Influence of the surface charge on the operation of ballistic T-branch junctions: a self-consistent model for Monte Carlo simulations, *Semicond. Sci. Technol.* 22, 663.
- Íñiguez-de-la-Torre I., Mateos J., Pardo D. & González T. (2008). Monte Carlo analysis of noise spectra in self-switching nanodiodes, *J. Appl. Phys.* 103, 024502, (2008)
- Íñiguez-de-la-Torre I., González T., Pardo D., Gardès C., Roelens Y., Bollaert S. & Mateos J. (2009a). Influence of the branches width on the nonlinear output characteristics of InAlAs/InGaAs-based three-terminal junctions, *J. Appl. Phys.* 105, 094504.
- Íñiguez-de-la-Torre I., Mateos J., Pardo D., Song A. M. & González T. (2009b). Noise and THz rectification linked by geometry in planar asymmetric nanodiodes, *Appl. Phys. Lett.* 94, 093512.
- Irie H., Diduck Q., Margala M., Sobolewski R. & M. J. Feldman (2008). Nonlinear characteristics of T-branch junctions: Transition from ballistic to diffusive regime, *Appl. Phys. Lett.* 93, 053502.
- Irie H. & Sobolewski R. (2010). Terahertz electrical response of nanoscale three-branch junctions, *J. Appl. Phys.* 107, 084315.
- Jacoboni C. & Lugli P. (1989). *The Monte Carlo method for semiconductor device simulation*, New York: Springer-Verlag.
- Kaushal V., Íñiguez-de-la-Torre I., Irie H., Guarino G., Donaldson W. R., Ampadu P., Sobolewski R. & Margala M. (2010). A study of geometry effects on the performance of ballistic deflection transistor, *IEEE Transactions on Nanotechnology*, in press, (published online DOI: 10.1109/TNANO.2010.2050069).
- Kim H., Min H. S., Tang T. W. & Park Y. J. (1991). An extended proof of the Ramo-Shockley theorem, *Solid-State Electron.* 34, 1251.
- Lai R., Mei X. B., Deal W.R., Yoshida W., Kim Y. M., Liu P.H., Lee J., Uyeda J., Radisic V., Lange M., Gaier T., Samoska L. & Fung A. (2007). Sub 50 nm InP HEMT Device with Fmax Greater than 1 THz, *IEDM Technical Digest*, 609.
- Landauer R. (1957). Spatial variation of currents and fields due to localized scatterers in metallic conduction, *IBM J. Res. Dev.* 1, 223.
- Lau B., Hartmann D., Worschech L. & and A. Forchel (2006). Cascaded Quantum Wires and Integrated Designs for Complex Logic Functions: Nanoelectronic Full Adder, *IEEE Trans. Electron Devices*, 53, 5, 1107.
- Lewén R., Maximov I., Shorubalko I., Samuelson L., Thylén L. & Xu H. Q. (2002). High frequency characterization of a GaInAs/InP electronic waveguide T-branch switch, *J. Appl. Phys.* 91, 2398.
- Mateos J., González T., Pardo D., Hoel V., Happy H. & Cappy A. (1999). Effect of the T-gate on the performance of recessed HEMT's. A MC analysis, *Semicond. Sci. Technol.* 14, 864.

- Mateos J., González T., Pardo D., Hoel V., Happy H. & Cappy A. (2000a). Improved MC algorithm for the simulation of  $\delta$ -doped AlInAs/GaInAs HEMTs, *IEEE Trans. Electron Devices* 47, 250.
- Mateos J., González T., Pardo D., Hoel V., Happy H. & Cappy A. (2000b). MC simulator for the design optimization of low-noise HEMTs, *IEEE Trans. Electron Devices* 47, 1950.
- Mateos J., Vasallo B. G., Pardo D., González T., Galloo J. S., Roelens Y., Bollaert S. & Cappy A. (2003a). Ballistic nanodevices for terahertz data processing: Monte Carlo simulations, *Nanotechnology* 14, 117.
- Mateos J., Vasallo B. G., Pardo D., González T., Galloo J. S., Bollaert S., Roelens Y. & Cappy A. (2003b). Microscopic modelling of nonlinear transport in ballistic nanodevices, *IEEE Trans. Electron Devices* 50, 1897.
- Mateos J., Vasallo B. G., Pardo D., González T., Pichonat E., Galloo J. S., Bollaert S., Roelens Y. & Cappy A. (2004). Non-linear effects in T-branch junctions, *IEEE Elec. Dev. Lett.* 25, 235.
- Mateos J., Vasallo B. G., Pardo D. & González T. (2005). Operation and high-frequency performance of nanoscale unipolar rectifying diodes, *Appl. Phys. Lett.* 86, 212103.
- Müller C. R., Worschech L., Höpfner P., Höfling S. & Forchel A. (2007). Monolithically integrated logic NOR gate based on GaAs/AlGaAs three-terminal junctions, *IEEE Electron Device Lett.* 28, 859.
- Palm T. & Thylen L. (1992). Analysis of an electron-wave Y-branch switch, *Appl. Phys. Lett.* 60, 2, 237-239.
- Palm T. & Thylen L. (1996). Designing logic functions using an electron waveguide Y-branch switch, *J. Appl. Phys.* 79, 8076.
- Rahman S. F. B. A., Nakata D., Shiratori Y. & Kasai S. (2009). Boolean logic gates utilizing GaAs three-branch nanowire junctions controlled by Schottky wrap gates, *Jpn. J. Appl. Phys.* 48, 06FD01.
- Rashmi, Bednarz L., Hackens B., Farhi G., Bayot V. & Huynen I. (2005). Nonlinear electron transport properties of InAlAs/InGaAs based Y-branch junctions for microwave rectification at room temperature, *Solid State Commun.* 134, 217.
- Reitzenstein S., Worschech L., Hartmann P. & Forchel A. (2002). Logic AND/NAND gates based on three-terminal ballistic junctions, *Electronics Lett.* 38, 951.
- Reitzenstein S., Worschech L. & Forchel A. (2004). Room temperature operation of an in-plane half-adder based on ballistic Y-junctions, *IEEE Electron Device Lett.* 25, 462.
- Rodilla H., González T., Pardo D. & Mateos J. (2009). High-mobility heterostructures based on InAs and InSb: A Monte Carlo study, *J. Appl. Phys.* 105, 113705.
- Sadi T., Dessenne F. & Thobel J. (2009). Three-dimensional Monte Carlo study of three-terminal junctions based on InGaAs/InAlAs heterostructures, *J. Appl. Phys.* 105, 053707.
- Shorubalko I., Xu H. Q., Maximov I., Omling P., Samuelson L. & Seifert W. (2001). Nonlinear operation of GaInAs/InP-based three-terminal ballistic junctions, *Appl. Phys. Lett.* 79, 1384.
- Shorubalko I., Xu H. Q., Maximov I., Nilsson D., Omling P., Samuelson L. & Seifert W. (2002). A novel frequency-multiplication device based on three-terminal ballistic junction, *IEEE Elec. Dev. Lett.* 23, 377.
- Song A. M., Lorke A., Kriele A., Kothaus J. P., Wegscheider W. & Bichler M. (1998). Nonlinear electron transport in an asymmetric microjunction: a ballistic rectifier, *Phys. Rev. Lett.* 80, 3831.

- Song A. M. (1999). Formalism of nonlinear transport in mesoscopic conductors, *Phys. Rev. B* 59, 9806.
- Song A. M., Omling P., Samuelson L., Seifert W., Shorubalko I. & Zirath H. (2001). Operation of InGaAs/InP-based ballistic rectifiers at room temperature and frequencies up to 50 GHz, *Jpn. J. Appl. Phys.* 40, L909.
- Song A. M., Missous M., Omling P., Peaker A. R., Samuelson L. & Seifert W. (2003). Unidirectional electron flow in a nanometer-scale semiconductor channel: A self-switching device, *Appl. Phys. Lett.* 83, 1881.
- Song A. M. (2004). *Room temperature ballistic nanodevices*, Encyclopedia of Nanoscience and Nanotechnology, 9, 371-389.
- Sun J., Wallin D., Brusheim P., Maximov I., Wang Z. G. & Xu H. Q. (2007). Frequency mixing and phase detection functionalities of three-terminal ballistic junctions, *Nanotechnology* 18, 195.
- Sun J., Wallin D., Maximov I. & Xu H. Q. (2008). A novel SR latch device realized by integration of three-terminal ballistic junctions in InGaAs/InP, *IEEE Electron Device Lett.* 29, 540.
- Suyatin D. B., Sun J., Fuhrer A., Wallin D., Fröberg L. E., Karlsson L. S., Maximov I., Wallenberg L. R., Samuelson L. & Xu H. Q. (2008). Electrical Properties of Self-Assembled Branched InAs Nanowire Junctions, *Nano Letters* 8, 1100
- Vasallo B. G., González T., Pardo D. & Mateos J. (2004). Monte Carlo analysis of four-terminal ballistic rectifiers, *Nanotechnology* 15, S250.
- Wesström J. O., (1999). Self-gating effect in the electron Y-branch switch, *Phys. Rev. Lett.* 82, 2564.
- Worschech L., Xu H. Q., Forchel A. & Samuelson L. (2001a). Bias-voltage-induced asymmetry in nanoelectronic Y branches, *Appl. Phys. Lett.* 79, 3287.
- Worschech L., Fischer F., Forchel A., Kamp M. & Schweizer H. (2001b). High frequency operation of nanoelectronic Y-branch at room temperature, *Jpn. J. Appl. Phys.* 40, L867.
- Worschech L., Schliemann A., Reitzenstein S., Hartmann P. & Forchel A. (2002). Microwave rectification in ballistic nanojunctions at room temperature, *Microelectron. Eng.* 63, 217.
- Worschech L., Reitzenstein S., Hartmann P., Kaiser S., Kamp M. & Forchel A. (2003). Self-switching of branched multiterminal junctions: A ballistic half-adder, *Appl. Phys. Lett.* 83, 2462.
- Xu H. Q. (2001). Electrical properties of three-terminal ballistic junctions, *Appl. Phys. Lett.* 78, 2064.
- Xu H. Q. (2002). A novel electrical property of three-terminal ballistic junctions and its applications to nanoelectronics, *Physica E* 13, 942-945.
- Xu H. Q., Shorubalko I., Wallin D., Maximov I., Omling P., Samuelson L. & Seifert W. (2004). Novel nanoelectronic triodes and logic devices with TBJs, *IEEE Electron Device Lett.* 25, 164.

# Estimation of Optical Properties in Postharvest and Processing Technology

László Baranyai  
*Corvinus University of Budapest*  
*Hungary*

## 1. Introduction

Non-destructive analysis and qualification methods are of great interest in agriculture, postharvest technology and food processing. Computer vision systems have the additional advantage that sensors do not touch the product and measurements can be performed from comfortable distance. Due to the recent developments in electronics, several commercial applications are already available for grading on the basis of visible attributes, near infrared (NIR) readings (GREEFA, The Netherlands; MAF Industries Inc., USA) and laser scattering (BEST NV, Belgium). Portable devices also exist providing optical quality assessment for flexible measurements and in vivo inspections (CP, Germany; Unitec S.p.A., Italy).

Information about interaction between light and biological tissue is essential in visual evaluation of fresh horticultural produces, raw materials and food, since optical signal is significantly affected by physical stage and valuable compounds of the tissue. Hidden physical damages in cucumber fruit, caused during harvest, transport and handling, were investigated in the spectral region of 950-1650 nm using hyperspectral imaging system (Ariana et al., 2006). Four specific wavelengths were selected for classification. The ratio of relative reflectances calculated as 988 nm to 1085 nm and the differences of this property obtained using 1346 nm and 1425 nm resulted in the highest classification rates. This study also confirmed that time plays very important role in qualification of perishable produces. Light transmittance through cucumbers in the range of 500-1000 nm was also investigated (Ariana & Lu, 2010). Transmittance for internally defected pieces and pickles was found to be generally higher compared to normal cucumbers. The wavebands around 745, 765, 885 and 965 nm were selected for best detection accuracy (94.7%). Advanced statistical methods, such as partial least square discriminant analysis (PLSDA) and k-nearest neighbor (KNN) may utilize the whole transmitted spectra. The highest classification rates of 97.3% and 88% were reached using PLSDA and KNN, respectively. Besides detection of damages and mechanical injury, estimation of quality parameters is also important for prediction of shelf-life and grading. Key quality attributes, mainly firmness and soluble solids content (SSC), were predicted for apple fruits based on multispectral imaging (Lu, 2004; Qing et al., 2007; 2008). Firmness and SSC were predicted for 'Red Delicious' apples, using the ratio of backscattering profiles in the wavelength range of 680-1060 nm, with the standard error of prediction (SEP) 5.8 N and 0.78%, respectively (Lu, 2004). Four specific wavelengths (680, 880, 905, and 940 nm) were selected for firmness and three (880, 905 and 940 nm) for SSC prediction. The size of the diffusively illuminated surface area and statistical description of the acquired

spatial intensity were used in prediction of firmness and SSC of 'Elstar' and 'Pinova' apples (Qing et al., 2007). Five selected wavelengths (680, 780, 880, 940 and 980 nm) resulted in SEP<13% in inter-cultivar validation. The distribution of measured surface intensity resulted in higher prediction accuracy than the size of the illuminated surface area. Additionally, the inter-cultivar validation on 'Pinova' apples confirmed the finding that fruit flesh firmness can be measured in parallel with fruit SSC. The results of this study also pointed out that surface curvature had significant effect on observed intensity. The apple fruit response to drought stress was also evaluated on 'Elstar' and 'Pinova' samples (Qing et al., 2008). Validation of the prediction model showed good agreement between predicted and reference values, SEP<10% for SSC and SEP<9% for firmness. According to the results, 12-13% effect of drought stress was observed compared to the sufficiently irrigated samples. Blue laser (408 nm) induced chlorophyll fluorescence was evaluated on 'Golden Delicious' apples (Noh & Lu, 2007). Fruit firmness, SSC, titratable acid content, skin and flesh color (chroma and hue) were predicted. Very good validation results were obtained for apple skin hue with the correlation coefficient of prediction of 0.94. Relatively good scores were obtained for fruit firmness, skin chroma, and flesh hue with correlation coefficients  $r \geq 0.74$ . The effect of processing technology on light-tissue interaction in banana slices was analyzed (Romano et.al, 2008). Three different drying temperatures were adjusted (53, 58 and 63°C) and a laser light source emitting at 670 nm was used. The gradient of the collected intensity profiles was observed to move closer to the incident point. Based on the extracted parameters, fresh and dry banana slices were distinguished with 100, 98 and 97.33% accuracy for 53, 58 and 63 °C, respectively. The estimation of drying time resulted in 78.19, 75.76 and 73.33% for 53, 58 and 63 °C, respectively. Tomato was also investigated using laser-scattering imaging (Tu et al., 2000). Low power (3 mW) laser module emitting at 670 nm was used to study the correlation between maturity, firmness and scattering. The total illuminated area correlated with acoustic firmness ( $r \approx 0.787$ ). This study also mention potential affecting factors such as fruit size and shape irregularities. This chapter would like to present practical applications of diffuse reflectance imaging technique, also called backscattering imaging, supported by Monte Carlo simulation. This type of simulation provides detailed information about the sensitivity of measured data to changes in fruit and vegetable tissue, environmental conditions, processing technology. The field of postharvest technology and process engineering requires sophisticated models taking several aspects into account, such as surface curvature, skin layer, measurement setup. Combination of experimental and simulation results help building models for prediction purposes.

## 2. Materials and methods

### 2.1 Effect of surface curvature

Reference methods and code libraries for Monte Carlo simulation (Wang et al., 1995) use planar surfaces and layers. Horticultural produces and foodstuff usually have more complex surface. In order to obtain realistic intensity values close to the in vivo readings, shape and location dependent correction is recommended during data post-processing. The Lambertian cosine correction is commonly applied for spherical shape and circular cross section (Kortüm, 1969; Qing et al., 2007). This approach assumes that camera is placed right above the sample and the vertical reflection is collected.

$$I_c = I_m \cdot \cos(a) = I_m \frac{\sqrt{R^2 - d^2}}{R} \quad (1)$$

Correction formula is explained by Fig. 1, where  $d$  means the observed distance between the selected location ( $I_c$ ) and incident point ( $IP$ ),  $R$  means the radius of the fruit,  $a$  means the central angle between incident and exit points,  $I_m$  is the intensity calculated for the specified location by simulation. According to the geometry, Equation 1 can be used to transform simulation result ( $I_m$ ) to acquired intensity ( $I_c$ ).

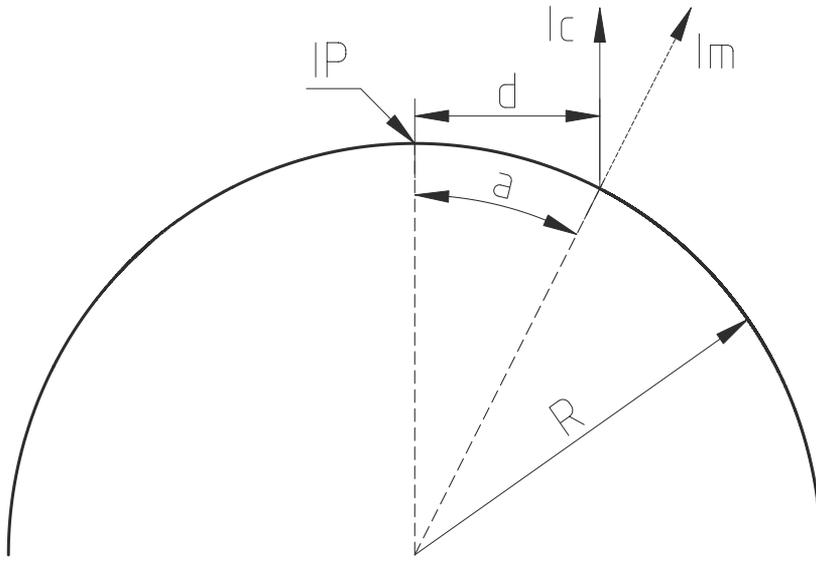


Fig. 1. Geometry for Lambertian cosine correction

More sophisticated correction formula was recently introduced on the basis of acceptance angle of the optical system (Lu, 2009; Lu & Peng, 2007; Peng & Lu, 2008). Figure 2 shows the main concept of this computation. The observed intensity may be calculated by integration of reflectance ( $R(r)$ ) over the acceptance angle ( $b - a$ ) of the zoom lens.

$$R(r) = \int_a^b I_m dS \cos^2 \Theta d\Theta = I_m dS \left[ \left( \frac{b}{2} + \frac{\sin 2b}{4} \right) - \left( \frac{a}{2} + \frac{\sin 2a}{4} \right) \right] \tag{2}$$

A correction factor can be used to normalize integration to the range of  $-c/2$  to  $c/2$ . Equation 3 shows the normalized calculation.

$$R(r) = \int_{-c/2}^{c/2} I_m dS \cos^2 \Theta d\Theta = I_m dS \left( \frac{c}{2} + \frac{\sin c}{2} \right) \tag{3}$$

**2.2 Incident light beam**

Light sources may be very different according to the applied system. Hyperspectral, multispectral or monochrome assembly require different hardware and wavelength ranges. Power regulation of lamps and laser diodes is essential. Halogen lamps are typically used in hyperspectral systems (Ariana & Lu, 2010) and their beam is focused on one point or line of the surface. Laser light diodes are typically used in multispectral systems (Baranyai & Zude, 2008; Qing et al., 2007; 2008). Laser beams could be also focused on one selected point or line

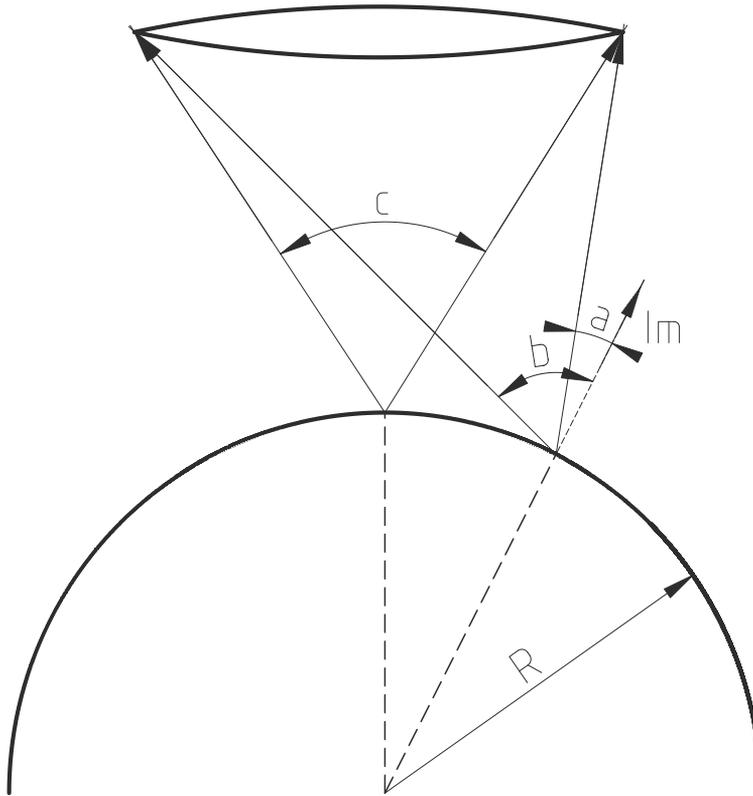


Fig. 2. Intensity correction using acceptance angle

but quality laser modules are collimated and the default size may be sufficient for practical purposes. The diameter of the incident light beam and the power distribution within its cross section affects the calculated launch position of the photon packages (Jacques, 1998; Wang et al., 1997). In the simplest case, uniform power distribution is assumed over the circular cross section. Equation 4 shows that a uniformly distributed random number ( $\xi$ ) can be expressed by an integral function based on the launch position ( $r$ ) and beam radius ( $b$ ).

$$\xi = \int_0^r p(r)dr = \int_0^r \frac{2\pi r}{\pi b^2} dr = \frac{r^2}{b^2} \quad (4)$$

Simulations usually assume that photon packages arrive perpendicular to the surface ( $0^\circ$  direction). Direction of the incident ray and the refractive index of the tissue affect the amount of photons reflected back from the surface. This reflection depends on the polarization of light (s- or p-polarized). Equation 5 shows the calculation of reflection coefficients based on the refractive indices ( $n_1$  and  $n_2$ ), incident angle ( $\Theta_i$ ) and travel direction in tissue ( $\Theta_t$ ).

$$R_s = \left( \frac{n_1 \cos \Theta_i - n_2 \cos \Theta_t}{n_1 \cos \Theta_i + n_2 \cos \Theta_t} \right)^2 \quad R_p = \left( \frac{n_1 \cos \Theta_t - n_2 \cos \Theta_i}{n_1 \cos \Theta_t + n_2 \cos \Theta_i} \right)^2 \quad (5)$$

Direct reflection from sample surface should be minimal, so that most of the photon packages can enter into the fruit tissue. Reflection coefficients for boundary of air and fruit tissue are

shown in Fig. 3. Type of polarization is usually unknown, especially due to the interaction between photon and biological tissue, therefore average value of  $R_s$  and  $R_p$  is commonly used (Wang et al., 1995). According to the close values of  $R_s$  and  $R_p$ , the range below  $20^\circ$  is recommended for incident light beam setup. Since the acquisition device is placed above the sample at  $0^\circ$  direction, the minimum incident angle should be also calculated to avoid direct reflection into the front lens.

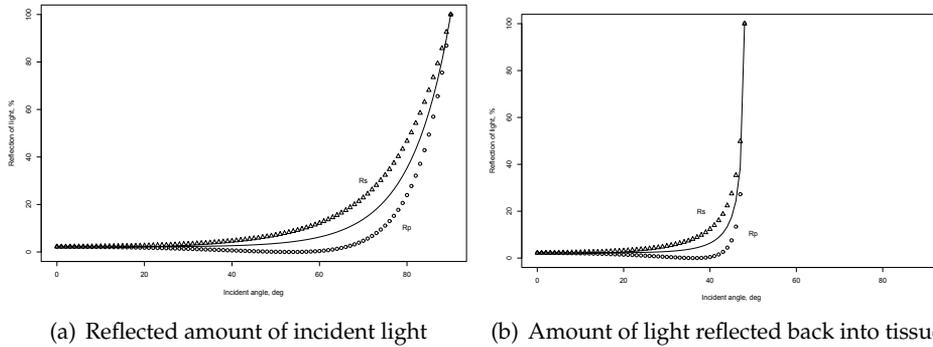


Fig. 3. Reflection of s-polarized ( $R_s$ ) and p-polarized ( $R_p$ ) photon packages entering (a) and leaving (b) fruit tissue of  $n=1.35$ . Average values are connected with line.

### 2.3 Evaluation of intensity profile

The observed intensity profile, also called backscattering profile, is calculated with radial averaging relative to the incident point. The typical logistic shape of this profile is analyzed in order to extract information about optical properties and quality. Due to the time consuming computation of Monte Carlo simulation, its primary role in agriculture and food science is to help construction and validation of inverse models used to estimate optical properties of biological tissues.

Intensity profiles can be evaluated with curve fitting of the diffusion theory model (Qin & Lu, 2008; 2009). Diffusion equation was validated for scattering dominant materials, where scattering coefficient is larger than absorption coefficient ( $\mu'_s \gg \mu_a$ ). Equation 6 shows the model used to describe diffuse reflectance ( $R_f(r)$ ) measured on the surface at  $r$  distance from incident point:

$$R_f(r) = \frac{a'}{4\pi} \left[ \frac{1}{\mu'_t} \left( \mu_e + \frac{1}{r_1} \right) \frac{\exp(-\mu_e r_1)}{r_1^2} + \left( \frac{1}{\mu'_t} + \frac{4A}{3\mu'_t} \right) \left( \mu_e + \frac{1}{r_2} \right) \frac{\exp(-\mu_e r_2)}{r_2^2} \right] \quad (6)$$

where  $\mu_a$  is the absorption coefficient,  $\mu'_s$  is the reduced scattering coefficient,  $a'$  is the transport albedo ( $a' = \mu'_s / (\mu_a + \mu'_s)$ ),  $\mu_e$  is the effective attenuation coefficient ( $\mu_e = [3\mu_a(\mu_a + \mu'_s)]^{1/2}$ ) and  $\mu'_t$  is the total attenuation coefficient ( $\mu'_t = \mu_a + \mu'_s$ ). Values of  $r_1$  and  $r_2$  are calculated using Equation 7:

$$r_1 = \left[ \left( \frac{1}{\mu'_t} \right)^2 + r^2 \right]^{1/2} \quad r_2 = \left[ \left( \frac{1}{\mu'_t} + \frac{4A}{3\mu'_t} \right)^2 + r^2 \right]^{1/2} \quad (7)$$

where  $A$  is the internal reflection coefficient of the tissue determined from refractive indices. This approach was successfully used to estimate the reduced scattering coefficient ( $\mu'_s$ )

and absorption coefficient ( $\mu_a$ ) for several types of fruits and vegetables (apple cultivars, cucumber, kiwifruit, peach, pear, plum, tomato, zucchini squash) within the wavelength range of 500-1000 nm (Qin & Lu, 2008).

Modified Lorentzian distribution functions (10 types) were tested for describing scattering intensity profile of 'Golden Delicious' apple fruit (Peng & Lu, 2008). The function introduced by Equation 8 was found to be the most appropriate for prediction of both firmness and SSC:

$$R = \frac{b}{1 + z/c^d} \quad (8)$$

where  $b$  is the peak value of the profile,  $z$  is the scattering distance,  $c$  is the full scattering width at half maximal peak value (FWHM) and  $d$  is the slope around FWHM. This function (Eq. 8) resulted in good prediction of apple fruit firmness with  $r=0.894$  and  $SEP=6.14$  N, and of SSC with  $r=0.883$  and  $SEP=0.73\%$ .

First order descriptive parameters, such as FWHM, slope and distance between incident and inflection points, were used to monitor changes in fruit tissue (apple, banana, kiwifruit) (Baranyai & Zude, 2008; 2009; Romano et.al, 2008). Sensitivity of the intensity profile to the changes in absorption, scattering coefficients and anisotropy factor ( $g$ ) was evaluated and anisotropy factor was found to be the most dominant parameter in case of apple and kiwifruit (Baranyai & Zude, 2008; 2009). The anisotropy factor was taken into account with the Heyney-Greenstein phase function (Jacques, 1998) instead of the approach of the reduced scattering coefficient ( $\mu'_s = [1 - g]\mu_s$ ). The effect of anisotropy factor appeared as rotation of the logarithmic profiles as it is shown on Fig. 4.

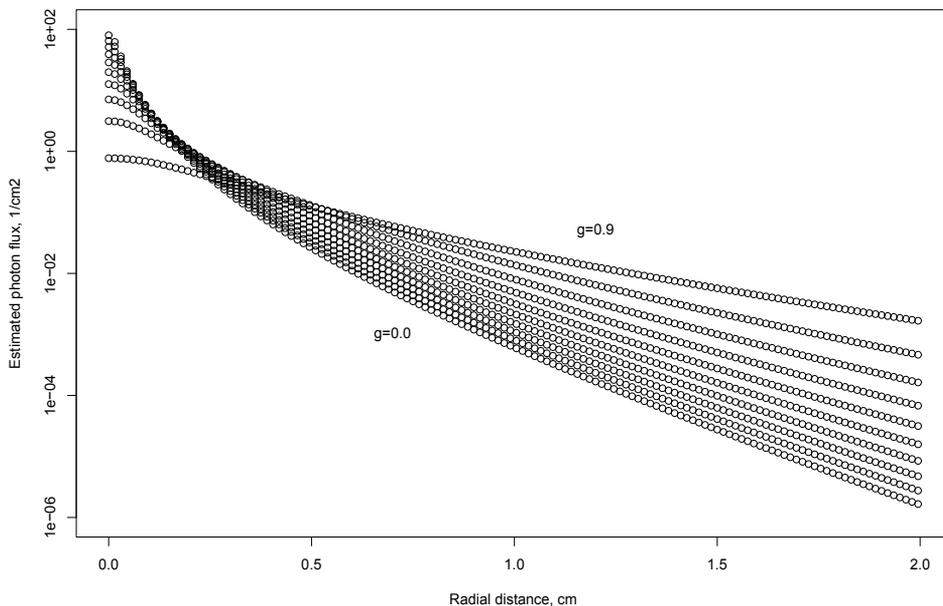


Fig. 4. Rotation of intensity profiles for apple ( $\mu_a=0.22$ ,  $\mu_s=30$ ,  $n=1.35$ )

The sensitivity was tested with ANOVA method. The statistical effect, Fisher score, of factors  $\mu_a$ ,  $\mu_s$  and  $g$  was also investigated in the range of  $g=0.5-0.9$  within  $\pm 10\%$  range around the selected values. Figure 5 shows the statistical effect of optical parameters on the shape of

intensity profile. The observed statistical power of optical properties exponentially increased with increasing value of anisotropy factor. Anisotropy factor was found to be the most important contributing attribute provided that  $g > 0.55$ .

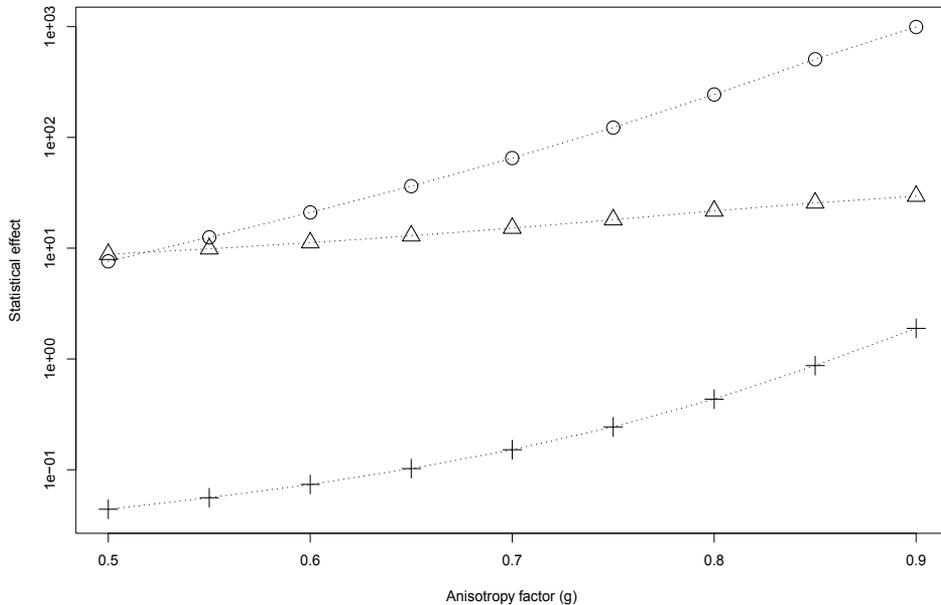


Fig. 5. Statistical effect of optical properties  $\mu_a$  (+),  $\mu_s$  ( $\Delta$ ) and  $g$  (o) on the shape of intensity profile

Based on the observed rotation, trigonometric function (Eq. 9) was selected to describe relationship between slope of the logarithmic profile ( $s$ ) and anisotropy factor ( $g$ ). Both determination coefficient ( $r^2=0.9996$ ) and Durbin-Watson autocorrelation test ( $D=2.1255$ ) results were the best for this type of trigonometric function among exponential and polynomial functions.

$$s = a + b \cdot \tan(g\pi/c) \quad (9)$$

The inverse function was successfully applied to estimate anisotropy factor for kiwifruits of different commercial grades and classify them based on this property. Significant difference ( $p < 0.01$ ) was found between commercial grades of ripe and overripe kiwifruit (Baranyai & Zude, 2009). However, commercial grade of unripe pieces overlapped others and statistical tests were unable to distinguish this class due to the high variance. The gradient was also found to change significantly for 'Idared' apples ( $p < 0.1$ ) and the backscattering area for 'Golden Delicious' apples ( $p < 0.05$ ) measured at 670 nm after bruising (Baranyai & Zude, 2008).

### 3. Case studies

#### 3.1 Controlled atmosphere cool storage of apples

Apple fruits (*Malus × domestica* 'Elstar' and 'Pinova') have been harvested in the orchard near Glindow (Germany). The middle of the field was located at latitude 52N 22' 14.96" and longitude 12E 52' 22.69". The selected area of 25 × 150 m had North-West to South-East

orientation and was split into upper and lower part according to the altitude. The upper part was suffering drought stress and both trees and fruits were obviously smaller. Harvested fruits were classified into the commercial grades of unripe, ripe and overripe according to the chlorophyll degradation (Fig. 6). Classes were well separated and no significant tendency was observed during storage. Harvested fruits were transferred into the storage facility immediately. Separate chambers were provided for apples of the same ripeness stage and cultivar. Temperature was adjusted to 2 °C. The atmosphere inside chambers consisted of 2% CO<sub>2</sub> and 1.5% O<sub>2</sub>. This controlled atmosphere cool storage started in August 2008 and took 157 and 164 days for 'Elstar' and 'Pinova', respectively. The continuous storage was broken for a few minutes in order to perform the measurements.

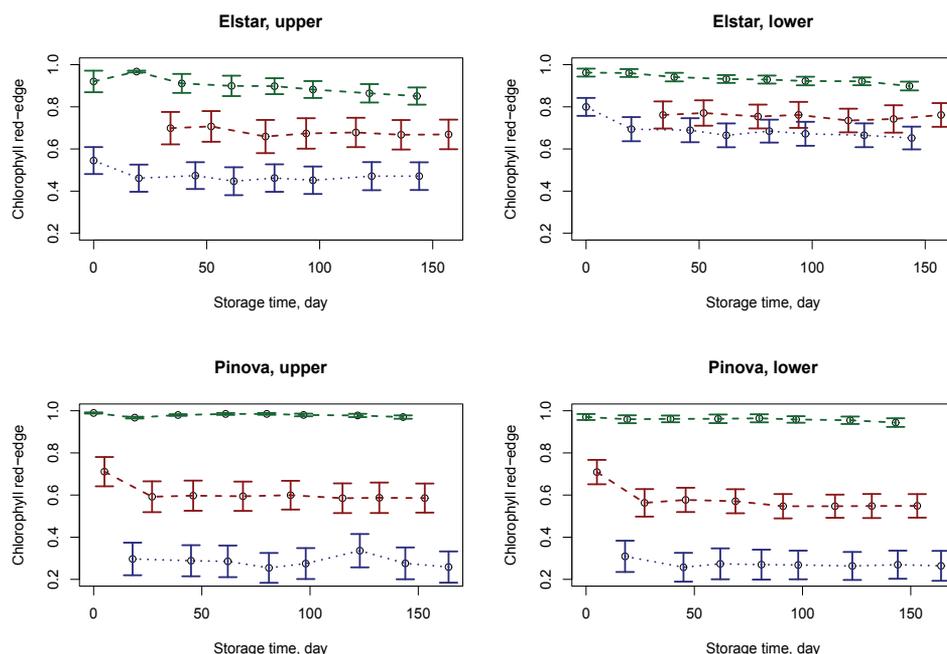


Fig. 6. Chlorophyll red-edge mean values of grades unripe (green), ripe (red) and overripe (blue) with 95% confidence intervals for apple cultivars (rows) and drought stress (column)

Digital images of 720×576 pixel size and 0.1694 mm/pixel resolution were acquired. Measurements took place in a darkroom in order to maximize signal to noise ratio. The vision system was consisted of a monochrome camera (JAI A50IR CCIR, JAI, Denmark), zoom lenses (model H6Z810, PENTAX Europe GmbH, Germany), external analog video converter (VRM AVC-1, Stemmer Imaging GmbH, Germany) and a laser module (LPM785-45C, Newport Corp., USA) emitting at 785 nm with 45 mW power. The laser module was aligned with 7° incident angle. The acquisition process was controlled by LabView 8.6 PDS software (National Instruments, USA) extended with a dynamic library of specific image processing functions. The optical parameters of simulation were adjusted in a wide range of  $\mu_a = 0.004\text{-}0.63\text{ cm}^{-1}$  and  $\mu'_s = 3.2\text{-}35\text{ cm}^{-1}$ . The size of the diffusively illuminated area at 50% peak intensity level showed good correlation with optical properties (Table 1). Reciprocal function fitted well to data points with  $r^2=0.998$ . The root mean square error of prediction (RMSEP) was  $0.37\text{ cm}^{-1}$  using 10% randomly selected data in 100 repetitions.

Correlation	$\mu_a$	$\mu'_s$	$\mu'_t$	$\mu_e$
Pearson (linear)	-0.0475	-0.8645	-0.8653	-0.6828
Spearman (rank)	-0.0393	-0.9990	-0.9996	-0.7354

Table 1. Correlation between optical parameters and FWHM

The gradient at the outline of the illuminated area did not change significantly during controlled atmosphere cool storage, estimated values for anisotropy factor (Eq. 9) changed in a narrow band, less, than 2.1%. This is in agreement with the observed chlorophyll red-edge readings presented in Figure 6.

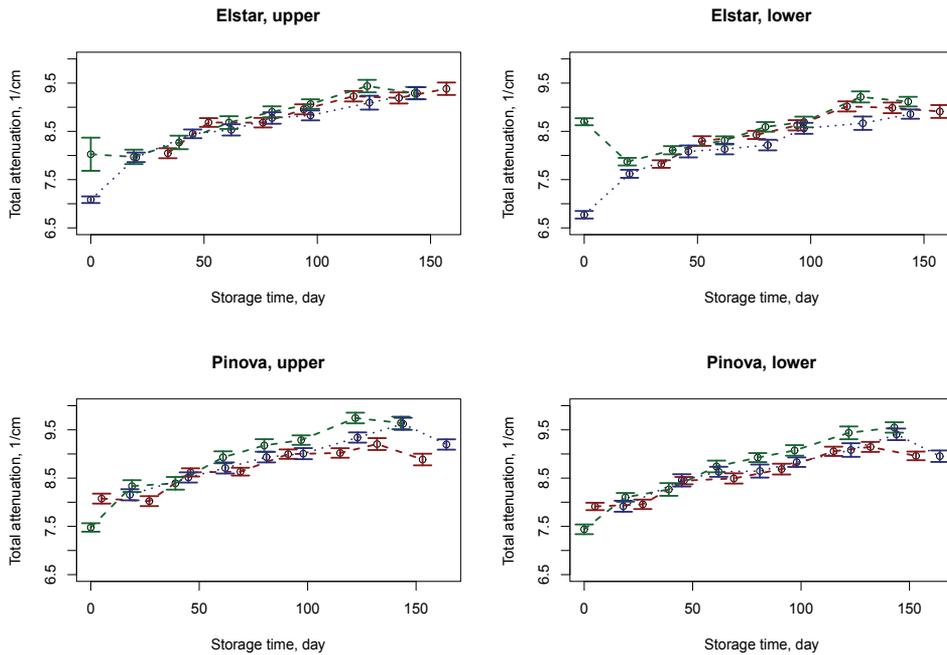


Fig. 7. Mean values of estimated total attenuation coefficient with 95% confidence intervals for apple cultivars (rows) and drought stress (column)

The diffusively illuminated area decreased significantly during storage, approximately 2.3% per month. Figure 7 presents changes of the estimated  $\mu'_t$  values during the experiment. The increasing value of this property results in decreasing value of the diffusion coefficient ( $D \sim 1/\mu'_t$ ). Decreasing value of diffusion coefficient together with the slightly changing anisotropy factor may support the assumption that light penetration depth can also decrease during storage of apple fruit.

### 3.2 Drying of banana slices

Cavendish bananas (*Musa × cavendishii* L.), originated in Central America, were used in drying experiment (Romano et.al, 2008). Sound pieces free from any visual defect were selected and stored for 24h at 13 °C and 82% relative humidity. Slices of 2.5-3.0 cm diameter and 1.0-1.3 cm thickness were prepared and placed on perforated steel trays. Three temperatures (53, 58 and 63 °C) and constant hot air velocity (0.75 m/s) were adjusted for 5h drying. The vision system was consisted of a laser diode emitting at 670 nm with

3 mW power (RS194-026, Global Laser Ltd., UK), a 3CCD digital camera (JVC KY-F50E, Victor Co., Japan) and grabber board (Optimas, Stemmer, Germany). The incident angle of  $15^\circ$  was applied. The collected intensity profiles were observed to shift closer to the incident point without significant change in the gradient. The color index measured with colorimeter (CR-300, Minolta, Japan) was found to increase until 4h drying in agreement with observed changes in intensity profile. The statistical analysis pointed out that changing moisture content had significant effect on observed profile ( $F=421.02$ ,  $p<2.2\times 10^{-16}$ ). It can be explained with reduced photon transport due to the increased optical density of shrunken tissue. Pearson correlation coefficient, calculated between backscattering area and moisture content, resulted in  $r^2=0.7214$ , 0.71 and 0.6698 for 53, 58 and 63  $^\circ\text{C}$ , respectively. Laser light distribution responded sensitively to changing tissue absorption caused by enhanced browning at higher temperature (63  $^\circ\text{C}$ ). Influence of colour changes was also investigated (Romano et.al, 2010). Different pre-treatments were applied to minimize colour degradation: chilling (4  $^\circ\text{C}$  for 18h), soaking in ascorbic and citric acid (1 min in a solution of 0.2% of 1:1 mixture), dipping in distilled water (30 s immersion). Untreated slices were separated as control set. The CIE  $L^*$  parameter was the most sensitive to colour change among treatments. Data analysis pointed out that higher moisture content resulted in deeper light propagation inside tissue and the diffusively illuminated area was affected only by the moisture content. Pre-treatments did not have significant influence on backscattering readings.

#### 4. Summary

Monte Carlo simulation was found to be an effective tool providing information on optical properties of fruit tissue and sensitivity of diffuse reflectance readings. Common simulation methods, such as MCML and CONV (Wang et al., 1995; 1997), should be changed slightly to build more accurate models compared to in vivo experiments. It was shown that surface curvature may affect observed intensity distribution and correction is required in post-processing. Direction of incident light beam seems to be optimal below  $20^\circ$  incident angle where minimal reflection is expected from surface. However, incident angle should be large enough to avoid direct reflection into the front lens.

Curve fitting of the diffusion equation and modified Lorentzian distribution functions was found to be successful in estimation of optical properties, firmness and SSC of various fruits and vegetables. Investigation of the shape of the backscattering profile revealed an additional potential parameter, the anisotropy factor ( $g$ ). Analysis of anisotropy factor found significant difference between commercial grades of kiwifruit and revealed effect of bruising in 'Idared' apples. Long term monitoring of apple storage and banana drying experiments proved that backscattering imaging can be utilized in postharvest and processing technology as well.

#### 5. Acknowledgements

The author wish to thank the Leibniz Institute of Agricultural Engineering Potsdam-Bornim e.V. (Germany) for help and the TÁMOP 4.2.1.B-09/1/KMR project for financial support.

#### 6. References

Ariana, D.P. & Lu, R. (2010). Hyperspectral waveband selection for internal defect detection of pickling cucumbers and whole pickles. *Computers and Electronics in Agriculture*, Vol. 74, No. 1, October 2010, 137-144, ISSN 0168-1699

- Ariana, D.P.; Lu, R.; Guyer, D.E. (2006). Near-infrared hyperspectral reflectance imaging for detection of bruises on pickling cucumbers. *Computers and Electronics in Agriculture*, Vol. 53, No. 1, April 2006, 60-70, ISSN 0168-1699
- Baranyai, L. & Zude, M. (2008). Analysis of laser light migration in apple tissue by Monte Carlo simulation. *Progress in Agricultural Engineering Sciences*, Vol. 4, No. 1, December 2008, 45-59, ISSN 1786-335X
- Baranyai, L. & Zude, M. (2009). Analysis of laser light propagation in kiwifruit using backscattering imaging and Monte Carlo simulation. *Computers and Electronics in Agriculture*, Vol. 69, No. 1, November 2009, 33-39, ISSN 0168-1699
- Jacques, S.L. (1998). Light distributions from point, line and plane sources for photochemical reactions and fluorescence in turbid biological tissues. *Photochemistry and Photobiology*, Vol. 67, No. 1, 23-32, ISSN
- Kortüm, G. (1969). *Reflectance spectroscopy. Principles, methods, applications*. Springer-Verlag, LCCCN: 79-86181
- Lu, R. (2004). Multispectral imaging for predicting firmness and soluble solids content of apple fruit. *Postharvest Biology and Technology*, Vol. 31, No. 2, February 2004, 147-157, ISSN 0925-5214
- Lu, R. (2009). Spectroscopic technique for measuring the texture of horticultural products: spatially resolved approach, In: *Optical monitoring of fresh and processed agricultural crops*, Zude, M. (Ed.) 391-423, CRC Press, ISBN 978-1-4200-5402-6, Boca Raton, USA.
- Lu, R. & Peng, Y. (2007). Development of a multispectral imaging prototype for real-time detection of apple fruit firmness. *Optical Engineering*, Vol. 46, No. 12, December 2007, 123201
- Noh, H.K. & Lu, R. (2007). Hyperspectral laser-induced fluorescence imaging for assessing apple fruit quality. *Postharvest Biology and Technology*, Vol. 43, No. 2, February 2007, 193-201, ISSN 0925-5214
- Peng, Y. & Lu, R. (2008). Analysis of spatially resolved hyperspectral scattering images for assessing apple fruit firmness and soluble solids content. *Postharvest Biology and Technology*, Vol. 48, No. 1, April 2008, 52-62, ISSN 0925-5214
- Qin, J. & Lu, R. (2008). Measurement of the optical properties of fruits and vegetables using spatially resolved hyperspectral diffuse reflectance imaging technique. *Postharvest Biology and Technology*, Vol. 49, No. 3, September 2008, 355-365, ISSN 0925-5214
- Qin, J. & Lu, R. (2009). Monte Carlo simulation for quantification of light transport features in apples. *Computers and Electronics in Agriculture*, Vol. 68, No. 1, August 2009, 44-51, ISSN 0168-1699
- Qing, Z.; Ji, B.; Zude, M. (2007). Predicting soluble solid content and firmness in apple fruit by means of laser light backscattering image analysis. *Journal of Food Engineering*, Vol. 82, No.1, September 2007, 58-67, ISSN 0260-8774
- Qing, Z.; Ji, B.; Zude, M. (2008). Non-destructive analyses of apple quality parameters by means of laser-induced light backscattering imaging. *Postharvest Biology and Technology*, Vol. 48, No. 2, May 2008, 215-222, ISSN 0925-5214
- Romano, G.; Argyropoulos, D.; Gottschalk, K.; Cerruto, E.; Müller, J. (2010). Influence of colour changes and moisture content during banana drying on laser backscattering. *International Journal of Agricultural and Biological Engineering*, Vol. 3, No. 2, June 2010, 46-51, ISSN 1934-6344
- Romano, G.; Baranyai, L.; Gottschalk, K.; Zude, M. (2008). An approach for monitoring the moisture content changes of drying banana slices with laser light backscattering

- imaging. *Food and Bioprocess Technology*, Vol. 1, No. 4, December 2008, 410-414, ISSN 1935-5130
- Tu, K.; Jancsó, P.; Nicolai, B.; De Baerdemaeker J. (2000). Use of laser-scattering imaging to study tomato-fruit quality in relation to acoustic and compression measurements. *International Journal of Food Science and Technology*, Vol. 35, No. 5, October 2000, 503-510, ISSN 1365-2621
- Wang, L.; Jacques, S.L.; Zheng, L. (1995). MCML - Monte Carlo modeling of light transport in multi-layered tissues. *Computer Methods and Programs in Biomedicine*, Vol. 47, No. 2, July 1995, 131-146, ISSN 0169-2607
- Wang, L.; Jacques, S.L.; Zheng, L. (1997). Conv - convolution for responses to a finite diameter photon beam incident on multi-layered tissues. *Computer Methods and Programs in Biomedicine*, Vol. 54, No. 3, November 1997, 141-150, ISSN 0169-2607

# MATLAB Programming of Polymerization Processes using Monte Carlo Techniques

Mamdouh A. Al-Harathi  
Chemical Engineering Department,  
King Fahd University of Petroleum & Minerals, Dhahran 31261,  
Saudi Arabia

## 1. Introduction

The expression "Monte Carlo method" is actually very general. Monte Carlo (MC) methods are stochastic techniques - meaning they are based on the use of random numbers and probability statistics to investigate problems. MC methods are used from economics to nuclear physics to regulating the flow of traffic. But the way MC methods are used, varies from one application to another. A plethora of algorithms are available for modeling a wide range of problems. But, to call something a "Monte Carlo" experiment, all you need to do is use random numbers to examine the problem. The use of MC methods to model physical problems allows us to examine more complex systems. For example, solving equations which describe the interactions between two atoms is fairly simple; solving the same equations for hundreds or thousands of atoms is impossible. With MC methods, a large system can be sampled in a number of random configurations, and that data can be used to describe the system as a whole (*Notes from University of Nebraska-Lincoln Physical Chemistry Lab*).

Free radical polymerization is one of the most widely used polymerization technique in polymer industry till date. Even a basic course on polymer science / engineering would without doubt start from free radical polymerization. Since it a pivot around which both academics and industrialist work around, it becomes imperative to study this class of polymerization. In this article, we would be illustrating how to model free radical copolymerization using the MC method based on Gillespie's algorithm. Even though a wide range of scholarly articles have been published on MC simulation in the field of polymerization from the 90's till the recent past (He et al., 1997,2000, Tobita 2001, 2003, Lu et a., 1993, Al-harathi et al., 2006,2007), but none of them give out the coding aspects of this process. Al -harathi et al. is one of the pioneers in modeling atom transfer radical polymerization, which is a subset/ branch of the free radical polymerization. Recently Al -harathi et al., have published two well accepted works on atom transfer radical copolymerization (Al -harathi et al., 2009 a & b). Combining these two aspects that the actual programming aspects have not been brought out till now and the authorial's immense experience in this field. We strongly feel the importance of educating chemical engineering students and industrials in this aspect. Even though we could have easily illustrated the homopolymerization and left the copolymerization to the readers, we on the contrary would like to educate them on a problem which is definitely difficult and hence by understanding these concepts, they can go on solve higher complex problems. We have dealt with atom transfer radical copolymerization directly which in fact is one step ahead of free radical

polymerization. Hence the reader can model both free radical and atom transfer radical copolymerization.

People might wonder why MC methods for modeling polymerization? Researchers have predominantly used population balances and method of moments for modeling polymerization. But its inability to predict the molecular weight distribution is a big drawback. MC method can easily predict the complete molecular weight distribution as shown in many referred journals (Al-harhi 2007). The ease of using MC method is another bigger advantage when compared to other numerical techniques. Hence it becomes necessary for students and researchers to understand this technique in a broader sense.

For the university student, this article provides a number of benefits:

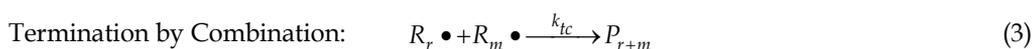
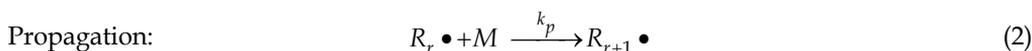
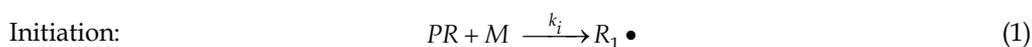
- Understanding the Gillespie's Algorithm
- A easy to understand procedure explained in detail
- The latest concept of atom transfer radical copolymerization which is not a subject matter in most general polymer science textbooks.

For the Industrialist, the article also provides a number of benefits:

- Our modeling technique produces a highly accurate result and hence of industrial importance
- Results like Molecular weight distribution, sequence length distribution, Polydispersity Index (PDI), conversion and other results can be easily modeled and relied upon on.
- Most of all, the ease with which the programming can be done.

## 2. General Monte Carlo procedure

The dynamic Monte Carlo approach used in this article is based on the method proposed by Gillespie (Gillespie 1977). Suppose we consider the following basic free radical kinetic scheme:



Where  $M$  is the monomer,  $PR$  is the propagating radical during initiation,  $R_r \bullet$  is polymer radical.  $P_r$  and  $P_m$  are the dead polymer chains.  $k_p$  is the propagation rate constant,  $k_{tc}$  is the rate constant of termination by combination,  $k_i$  is the initiation rate constant, and the subscripts  $r$  and  $m$  indicate the number of monomer molecules in the chain.

First the deterministic, or experimental, rate constant ( $k^{exp}$ ) should be changed to stochastic, or Monte Carlo, rate constants ( $k^{MC}$ ) according to the following equations

$$k^{MC} = k^{exp} \quad \text{for first order reactions} \quad (4)$$

$$k^{MC} = \frac{k^{exp}}{VN} \quad \text{for bimolecular reactions between different species} \quad (5)$$

$$k^{MC} = \frac{2k^{exp}}{VN} \quad \text{for bimolecular reactions between similar species} \quad (6)$$

Secondly all concentrations should be transformed to number of molecules in the control (simulation) volume  $V$ ; in our example we have only monomer concentration, consequently:

$$X_m = [M] NV \quad (7)$$

Where  $N$  is the avagadro's number

Then calculate the reaction rate for every reaction according to the equations:

Rate of initiation:

$$R_i = k_i^{MC} X_{pr} X_m \quad (8)$$

Rate of propagation:

$$R_p = k_p^{MC} X_r X_m \quad (9)$$

Rate of termination by combination:

$$R_{tc} = \frac{k_{tc}^{MC} X_r (X_r - 1)}{4} \quad (10)$$

Where  $X_r$  and  $X_m$  are the number of polymer radicals and monomer molecules respectively.

The total reaction rate ( $R_{sum}$ ) is then calculated as the summation of the individual reaction rates.

Then the probability of any reaction ( $P_v$ ) taking place at a given time is calculated by the following equation

$$P_v = \frac{R_v}{R_{sum}} \quad (11)$$

Then the following relation is used to determine which reaction type will take place at a given polymerization time

$$\sum_{v=1}^{\mu-1} P_v < r_1 < \sum_{v=1}^{\mu} P_v \quad (12)$$

where  $\mu$  is the number of the selected reaction type and  $r_1$  is a random number uniformly distributed between  $[0, 1]$ . Another random number is generated to determine the time interval ( $\tau$ ) between two consecutive reactions. The time step is related to the inverse of total stochastic rates and the natural logarithmic of  $r_2$  according to the equation:

$$\tau = \frac{1}{\sum_{v=1}^N R_v} \ln \left( \frac{1}{r_2} \right) \quad (13)$$

The algorithm for a general Monte Carlo simulation is of the following steps:

1. Input the deterministic reaction rate constant,  $k_1^{\text{exp}}$ ,  $k_2^{\text{exp}}$ ,  $k_3^{\text{exp}}$ , ...  $k_{\mu}^{\text{exp}}$ , simulation volume,  $V$ , avagadro's number,  $N$ , reactant concentration (mol/Volume)
2. Set time to zero and conversion to zero,  $t=0$  and  $x=0$
3. Calculate the stochastic rate constants,  $k_1^{MC}$ ,  $k_2^{MC}$ , .....  $k_{\mu}^{MC}$  using equation 4, 5 or 6
4. Calculate and store the rates of reaction  $R_1$ ,  $R_2$  .....  $R_{\mu}$  for the selected reaction mechanism

5. Calculate and store the sum of the rates of reaction,  $R_{sum}$  according to the following equation:  $R_{sum} = \sum_{\mu=1}^M R_{\mu}$  where  $R_{\mu}$  is the rate of the  $\mu^{\text{th}}$  reaction and  $M$  is the number of reactions.
6. Get two random numbers,  $r_1$  and  $r_2$ , uniformly generated between 0 and 1, and calculate  $\mu$  and  $\tau$  according to equations 12 and 13.
7. Select which reaction will occur according to equation 12
8. Update the number of molecules of each type in the reactor
9. Update the simulation time,  $t$ , by  $t = t + \tau$ , and calculate the conversion,  $x$
10. Return to step 4 until we obtain the desired polymerization time  $t = t_{\text{final}}$  or final conversion  $x = x_{\text{final}}$

The sequence of the steps could be altered or changed depending on convenience, flexibility and programming skills of the programmer.

### 3. Detailed Monte Carlo procedure and MATLAB programming:

Before moving to the modeling aspect, a brief introduction of atom transfer radical polymerization would give a better understanding and knowledge on this much researched topic. It was in the year 1995, discoveries on this process was made by Matyjaszewski's group and Sawamoto's group. Till date several reviews, books, and book chapters have summarize hundreds of papers that appeared in the literature on ATRP of a large variety of monomers (Matyjaszewski 2006). ATRP can synthesize various polymers with controlled molecular weight and narrow MWD. It can be carried out in a wide range of polymerization temperatures and is not very sensitive to the presence of oxygen and other inhibitors.

Figure 1 shows the general mechanism of ATRP. In addition to the monomer, the ATRP system consists of an initiator that has an easily transferable halide atom (RX) and a catalyst.

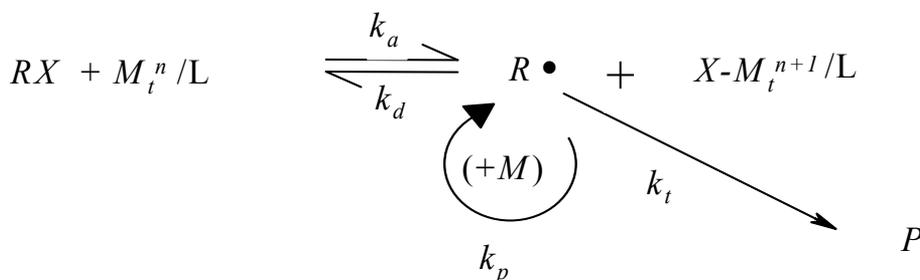


Fig. 1. ATRP Mechanism. RX: dormant species (alkyl halide);  $M_t^n / L$ : activator (metal complex);  $R \cdot$ : propagating radical;  $X-M_t^{n+1} / L$ : deactivator; M: monomer; P: dead chain.

The catalyst (or activator) is a lower oxidation state metal halide ( $M_tX$ ) with a suitable ligand (L). Polymerization starts when the halide atom transfers from the initiator to the catalyst to form a free radical and a higher oxidation state metal halide  $M_t^{n+1}X$  (deactivator). This step is called activation or forward reaction. The deactivation step or backward reaction pushes the reaction to form dormant species (RX) rather than the radical's. The reaction of monomer molecules (M) in the propagation step is similar to conventional free radical polymerization.

For the modeling, we consider the following mechanism for atom transfer radical copolymerization. By removing equation 14 and 15 and by adding the dissociation of the

initiator to give the free radicals would transform the process to free radical polymerization. Hence without any doubt we can apply our procedure to both free radical copolymerization and atom transfer radical copolymerization. Even though a general mechanism of Monte Carlo approach has been given before, a detailed description of how Monte Carlo simulations were carried for a specific polymerization process like atom transfer radical copolymerization would be interesting and will help the reader understand more about this process.

Initiation steps:



Equilibrium and propagation steps:

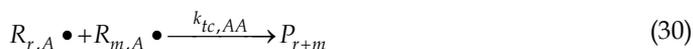


Transfer to monomer steps:

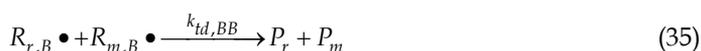
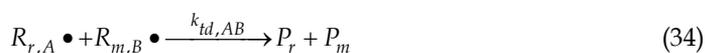
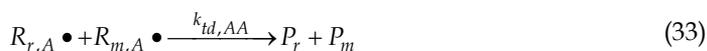




Termination by combination:



Termination by disproportionation:



In Equations (14) to (35),  $I$  is the initiator,  $C$  and  $CX$  are the catalyst in its low and high valence states,  $M_A$  and  $M_B$  are the comonomers,  $R_{r,A} \bullet$  and  $R_{r,B} \bullet$  are polymer radicals terminated in monomer  $A$  and  $B$ ,  $P_r$  is a dead polymer chain,  $D_r$  is a dormant polymer chain,  $k_i$  is the initiation rate constant,  $k_a$  is the activation rate constant,  $k_d$  is the deactivation rate constant,  $k_p$  is the propagation rate constant,  $k_{tc}$  is the rate constant of termination by combination,  $k_{td}$  is the rate constant of termination by disproportionation,  $k_{tr}$  is the transfer rate constant, and the subscripts  $r$  and  $m$  indicate the number of monomer molecules in the chain. The subscript  $A$  denotes that the chain ends with monomer  $A$  and the subscript  $B$  has an equivalent meaning.

The following assumptions and hypothesis are made in this mechanism:

1. All reactions are irreversible
2. All reactions are elementary
3. Rate Constants are chain length independent
4. Initiator and catalyst efficiencies are constant
5. Thermal initiation does not occur

Step 1:

As stated before we input the following constant:

- Temperature
- Concentration of Monomer  $A$  ( $CMA$ ) and Monomer  $B$  ( $CMB$ )
- Avogadro's Number ( $N$ )
- Experimental Rate Constants ( $k_1^{exp}, k_2^{exp}, k_3^{exp}, \dots, k_\mu^{exp}$ )
- Reactivity ratios  $r_1$  and  $r_2$

Step 2:

During the process of polymerization in ATRcP, five different species are formed within the reactor namely:

- Dormant chain with end group corresponding to monomer A ( $D_{r,A}$ )
- Dormant chain with end group corresponding to monomer B ( $D_{r,B}$ )
- Growing Polymer Radicals with end group corresponding to monomer A ( $R_{r,A}$ )
- Growing Polymer Radicals with end group corresponding to monomer B ( $R_{r,B}$ )
- Dead Polymer chains (P)

MATLAB heavily relies on the creation of matrices in order to transform physical processes and to simulate such processes. A normal matrix represented by [ ] needs to have equal number of rows when constructing a column matrix.

Example:  $A = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}$  is the correct representation for a normal matrix

But in the case of polymerization we have thousands of unequal polymer chains and hence to represent such a scenario, a special type of matrix called cell matrix represented by { } have been utilized in our case.

Example:  $B = \left\{ \begin{array}{c} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \\ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} \\ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \end{array} \right\}$  is a special type of a collection of independent matrices.

Such a matrix allows us to know the following vital information:

1. Total number of chains, corresponding to 4 in example B
2. Chain length of independent chains - corresponding to (5,3,8,2) from example B
3. Sequence Length Distribution

Consider chain 3(the longest chain), the program can access each element within the chain. If we assign 1 for monomer A and 0 for monomer B, the program scans for the sequence and gives us the following diads.

[10]-AB, [00]-BB, [01]-BA, [01]-BA. Thus by assigning 0 and 1 for monomer A and B, we can easily know the diads and triads since we are able to access every element in every chain.

The sequence can be obtained using the following syntax in MATLAB:

```
Diads(0,0) = length(findstr(CELLMATRIXCONVERTEDTONORMALMATRIX, [0,0]))
```

A "cell2mat" command and assigned to a variable "vec"

Note that you have to convert the cell matrix to a normal matrix and assign it another variable. We would see how to perform this operation and what is the term "vec" in detail in the coming sections. Figure 1 shows the diad prediction from our simulations. We have chosen the Styrene Methyl methacrylate( St-MMA) copolymer due to its applications as commodity plastics and also as engineering polymer. We have discussed the entire simulation results and tabulated the rate constants used, in our technical paper (Al-harathi et al 2009 a, b).

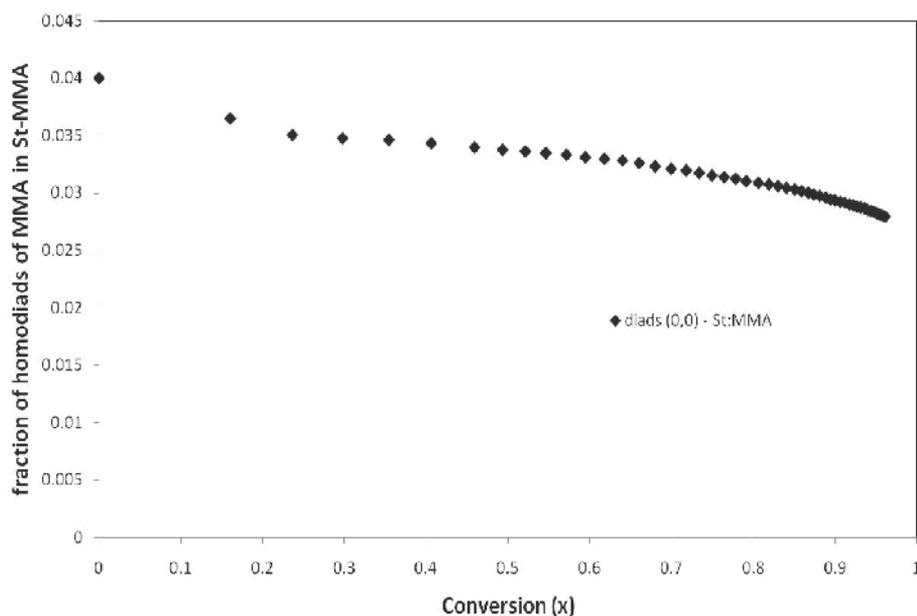


Fig. 1. Monte Carlo simulation results for the copolymerization of styrene and methyl methacrylate: fraction of homodiad of MMA as a function of conversion. The initial comonomer molar fractions in the reactor were  $f_{0,St} = 0.75$ ,  $f_{0,MMA} = 0.25$ . The rate constants and other parameters are tabulated in our journal paper (Al-harthi 2009a)

#### 4. Molecular Weight Distribution(MWD)

Chain lengths are obtained from individual chains. From the chain length, the molecular weight could be obtained by knowing the physical properties of the monomers. Since individual chain lengths and their corresponding molecular weights are obtained, the MWD could be generated using Monte Carlo simulation.

Another example is Figure 2 which shows the strength of Monte Carlo simulations to predict the molecular weight distribution.

#### 5. The type of end group in the chain ( monomer A ended or monomer B)

Thus we allocate four cell matrices to  $D_{r,A}$ ,  $D_{r,B}$ ,  $R_{r,A}$  and  $R_{r,B}$  to the following cell matrices Da, Db, Ra and Rb respectively. The dead polymer chains are denoted by a normal matrix P. The dead polymers chains are classified as a normal matrix since only addition of terminated chains are involved without any further operations on them.

Step 3:

The number of monomer molecules, catalyst and initiator are calculated using the following relationship.

$$X_A = [M_A]NV$$

$$X_B = [M_B]NV$$

$$X_i = [I]NV$$

$$X_C = [C]NV$$

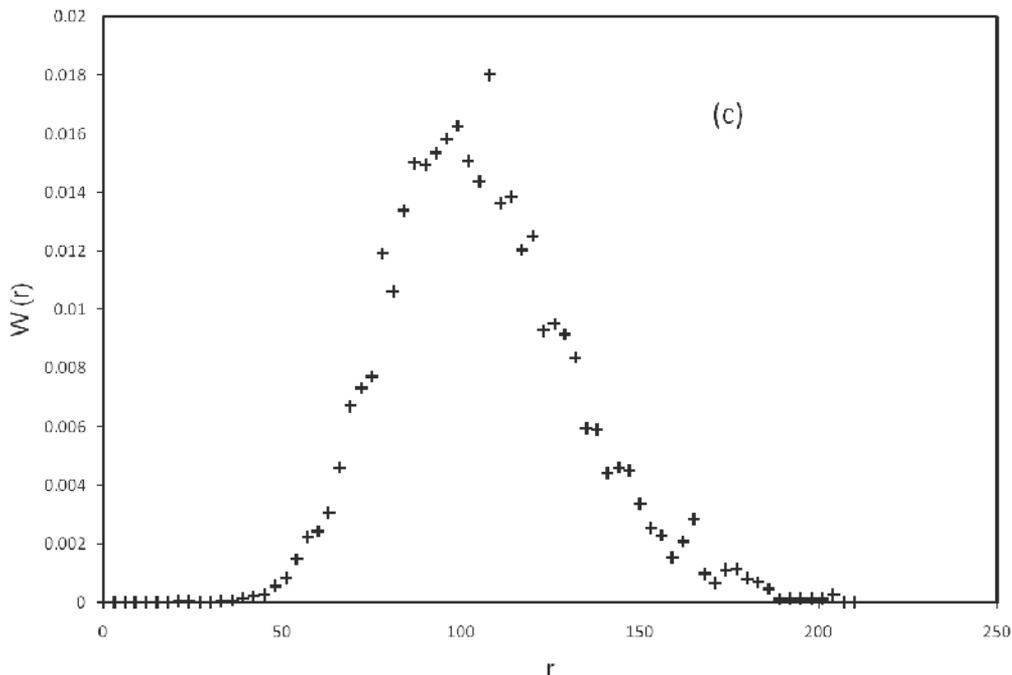


Fig. 2. Molecular weight distribution of styrene methyl methacrylate copolymers at a conversion of 0.99. The initial comonomer molar fractions in the reactor were  $f_{0,St} = 0.5$ ,  $f_{0,MMA} = 0.5$ . The rate constants and other parameters are tabulated in our journal paper (Alharthi 2009a)

The objective is to eliminate the volume in the concentration and find the number of molecules of monomer, catalyst and initiator.

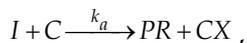
The experimental rates are also converted into stochastic rates by equations 4, 5 and 6.

Step 4:

The rate of the reaction is calculated and is explained by the following examples.

Example 1: Initiation

Consider the following initiation reaction (equation 14)



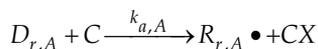
The rate of the reaction would be calculated as

$$R(1) = k_{a1}^{MC} X_C X_i$$

Where  $k_{a1}^{MC}$  is the stochastic rate constant of activation for monomer A.

Example 2: Equilibrium

Consider the following equilibrium reaction (equation 18)



The rate of the reaction would be calculated as

$$R(5) = k_{a1}^{MC} X_{da} X_C$$

Thus all the rate of reaction are calculated as a product of their respected stochastic rate and number of molecules of either dormant chains, catalyst, monomers or growing polymer radical chains

Step 5:

The next step is to calculate the total rate and probability of each reaction. Summation of all the rates gives us the total rate of reaction. The probability of a particular reaction to take place is calculated from equation 2.13

Example 3:

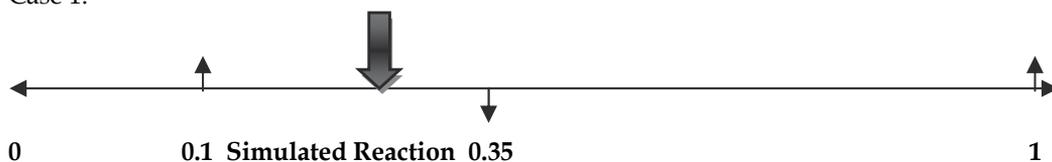
$$\text{probability of activation reaction} = \frac{\text{Rate of activation reaction}}{\text{Total rate of reaction } (R_{sum})}$$

While the probability of the specific reaction type is obtained by generating a random number uniformly distributed between 0 and 1. Using the random number and the probability of individual reactions, a suitable reaction is chosen. This is illustrated with an example below

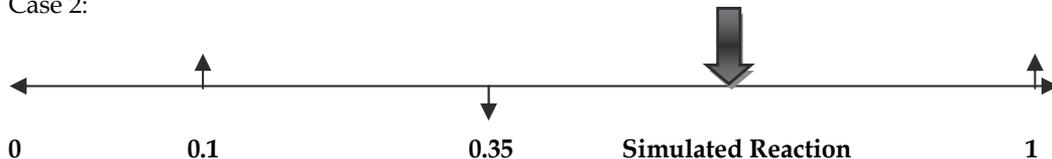
Example 4:

A random number is generated in MATLAB by the command rand ( $r_1 = \text{rand}$ ) results in an output of 0.25 (Case 1). Consider that individual probabilities of three reactions are  $P_1 = 0.1$ ,  $P_2 = 0.25$ ,  $P_3 = 0.65$ . The program would choose the reaction having the probability of 0.25 ( $P_2$  in this case which refers R (2) for the reaction type). Consider another case where we again generate a random number and we get an output of 0.5 (Case 2). In this case the program will pick a value which is above  $P_2$  which is  $P_3$ . Thus the program chooses which reaction type to take place at every loop.

Case 1:



Case 2:



Step 6:

Calculation of time

Another random number ( $r_2$ ) is chosen between 0 and 1 and used to calculate the time step by the following equation

$$\tau = \frac{1}{R_{sum}} \ln \left( \frac{1}{r_2} \right)$$

The time (t) is updated at every loop with the following relation:

$$t = t + \tau$$

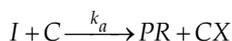
Step 7:

Simulation of selected reaction

Once a particular reaction is chosen randomly as is in the case of a polymerization reactor, the next step is in carrying out the selected reaction by using the language of MATLAB. Many examples are given in order to clearly illustrate the procedure adopted while handling different type of reactions like initiation, equilibrium, propagation, cross propagation, transfer, cross transfer and bimolecular termination.

Example 5: Initiation

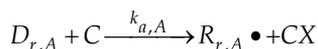
Consider the following reaction



In this reaction, the initiator and catalyst lose molecules while an increase in the propagating radical and the deactivator/catalyst in the lower valence take is observed. Since Monte Carlo simulation works at the molecular level, we increase and decrease the number of molecules according to the reaction. Thus in this type of reaction we perform the following operation

- Decrease the number of Initiator molecules by 1:  $X_i = X_i - 1$
- Decrease the number of Catalyst molecules by 1:  $X_c = X_c - 1$
- Increase the number of the PR molecules by 1:  $X_{pr} = X_{pr} + 1$
- Increase the number of CX molecules by 1:  $X_{cx} = X_{cx} + 1$

Example 6: Equilibrium



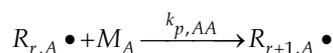
As seen from the reaction, there is a decrease in the catalyst and dormant chains while an increase in polymer radical and the lower valence catalyst. We should note that the dormant and the growing polymer radicals are in form of chains and we have represented them as cell matrices earlier. Hence a more sophisticated approach has to be adopted than the one adopted during initiation.

The following operations are performed:

- Decrease the number of catalyst molecules by 1:  $X_c = X_c - 1$
- We need to add one molecule to  $X_{ra}$  (number of growing radical chain) and reduce one molecule from  $X_{da}$  (number of dormant chains) .  
 $X_{ra} = X_{ra} + 1;$   
 $X_{da} = X_{da} - 1;$
- We need to randomly select one dormant chain from the cell matrix Da and transfer it to the cell Ra. To perform this physical phenomenon in MATLAB, we use the following procedure.
  - Find the total number of chains of the Dormant Radicals :  
 $>> \text{LEN} = \text{length}(\text{Da})$

- Randomly choose one chain among the available chains:  
`>> rvec = randsample(LEN,1)`
  - The randomly chosen chain (rvec) would be placed in the growing polymer radical as follows:  
`>> Ra[Xra] = Da{rvec}`
  - This ensures that the chain is placed in the new location created before and hence avoiding overlaying of chains.
- Finally we need to remove the randomly chosen chain (rvec) from Da since it has been placed in Ra. The following command removes the chosen chain from Da.  
`>> Da(rvec) = [ ]`

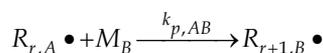
Example 7: Propagation



The following example shows how propagation reactions are handled in the program. We have a decrease in monomer A and an increase in the chain length of Ra. Hence the following operations are performed.

- Decrease in the number of monomer A by 1 :  $X_A = X_A - 1$
- The physical phenomenon is the addition of monomer of the same kind (Monomer A to Radical A) to a particular chain in Ra which increases the chain length by adding one Monomer A unit. To incorporate this physical phenomena in the program the following operation was done:
  - Find the total number of Radical chains :  
`>> LEN = length (Ra)`
  - Randomly choose one chain :  
`>> rvec = randsample (LEN,1)`  
 Convert the obtained chain which is a cell matrix type to a normal matrix in order to perform further operation on it since a cell matrix has access to limited basic operation. Hence the need arises to convert it to a normal matrix at certain junctures.  
`>> vec = cell2mat (Ra (rvec))`  
 is the command for such an operation
  - Adding the monomer to the particular chain:  
`>> lvec = length (vec) +1`
  - Specifying the type of monomer added (Whether monomer A or B). Monomer A has a indexing of "1" and monomer B has an indexing of "0" in the program. Thus the following assigns that the monomer added is of type A.  
`>> vec (lvec) =1`  
 performs the indexing operation.
  - Finally we need to have the chain in the cell matrix form due to aforementioned advantages of using a cell matrix. Hence the final part of the program is to convert the normal matrix to a cell matrix and place it in Ra  
`>> Ra(rvec) = vec`

Example 8: Cross propagation



The programming of cross propagation is slightly different from the previous section since two dissimilar radicals are involved. As emphasized before here a decrease in monomer B and Ra is seen while an increase is observed in Rb.

- Decrease in the number of monomer B by 1 :  $X_B = X_B - 1$
- Increase in the number of growing polymer radical (Rb) :  $X_{rb} = X_{rb} + 1$
- The physical phenomenon is the addition of monomer of the different kind (Monomer B to Radical A) while creating an increase in the chain length of Rb by adding one Monomer B unit. To incorporate this physical phenomena in the program the following operation was done:
  - Find the total number of Radical chains :  
`>> LEN = length (Ra)`
  - Randomly choose one chain :  
`>> rvec = randsample (LEN,1)`  
Convert the obtained chain which is a cell matrix type to a normal matrix in order to perform further operation on it since a cell matrix has access to limited basic operation. Hence the need arises to convert it to a normal matrix at certain junctures.  
`>> vec = cell2mat (Ra (rvec))`  
is the command for such an operation
  - Adding the monomer to the particular chain:  
`>> lvec = length (vec) +1`
  - Specifying the type of monomer added (Whether monomer A or B). Monomer A has a indexing of "1" and monomer B has an indexing of "0" in the program. Thus the following assigns that the monomer added is of type A.  
`>> vec (lvec) = 0` performs the indexing operation.
  - Finally we need to have the chain in the cell matrix form due to aforementioned advantages of using a cell matrix. Hence the final part of the program is to convert the normal matrix to a cell matrix and place it in Ra  
`>> Rb (xrb) = vec`
- Finally the chain which was taken from Ra has to be removed since it has been transferred to Rb.  
`Xra = Xra - 1`  
`>> Ra (rvec) = []` removes the chain which had been transferred to Rb.

The technique for handling transfer and cross transfer reactions are similar to the propagation reactions and hence will not be discussed.

A few of the results at this stage are given which could be used as a starting point and validation check for the results obtained using our methodology. Conversion is very well related to the propagation reaction and hence we would like to show a sample result for the linear relation obtained between conversion and the average chain length. Figure 3 shows the linear relationship between the average chain length and conversion.

Example 9: Bimolecular termination



Termination is the process by which two growing polymer radical chains combine and form a dead polymer. The programming is done in the following manner

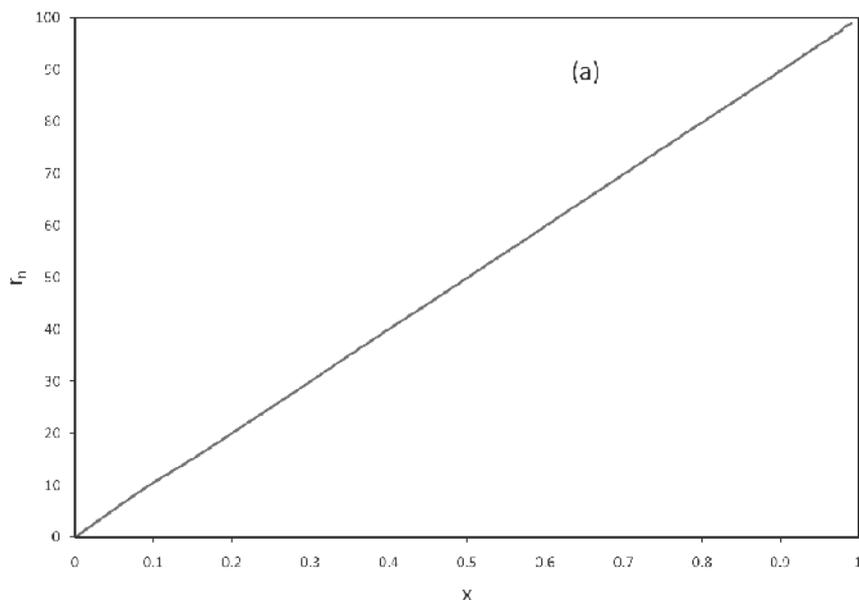


Fig. 3. Monte Carlo simulation results for the copolymerization of styrene and methyl methacrylate: average chain length as a function of conversion. The initial comonomer molar fractions in the reactor were  $f_{0,St} = 0.5$ ,  $f_{0,MMA} = 0.5$ . The rate constants and other parameters are tabulated in our journal paper (Al-harathi 2009a)

- According to the equation, two chains from  $R_a$  will combine to form dead polymer and hence  $R_a$  should have a minimum of two chains. This check is done as follows  
`>> LEN = length(Ra)` and only if  $LEN > 2$ , the reaction would proceed to completion.
- Two random chains are chosen from  $R_a$  as follows:  
`>> rvec1 = randsample(LEN, 1)`  
`>> rvec2 = randsample(LEN, 1)`
- The following chains are in the cell matrix form and hence need to be converted to a normal matrix in order to perform further operation on them, hence the following code is utilized:  
`>> vec1=cell2mat(Ra(rvec1))`  
`>> vec2=cell2mat(Ra(rvec2))`
- The number or the length of each chain is calculated as follows:  
`>> lvec1=length(vec1)`  
`>> lvec2=length(vec2)`
- We have the formation of a dead polymer chain and hence an increase in the number of dead polymer takes place as follows:  
 $X_p = X_p + 1$
- Now the summation of the two growing radical chains takes place becoming a dead polymer as follows  
`>> P(xp)= lvec1+lvec2`
- Finally the growing chains which formed the dead polymer should be removed from  $R_a$  as follows;  
 $X_{ra} = X_{ra} - 2$ ;  
`>> Ra{rvec1}=[];`  
`>> Ra{rvec2}=[];`

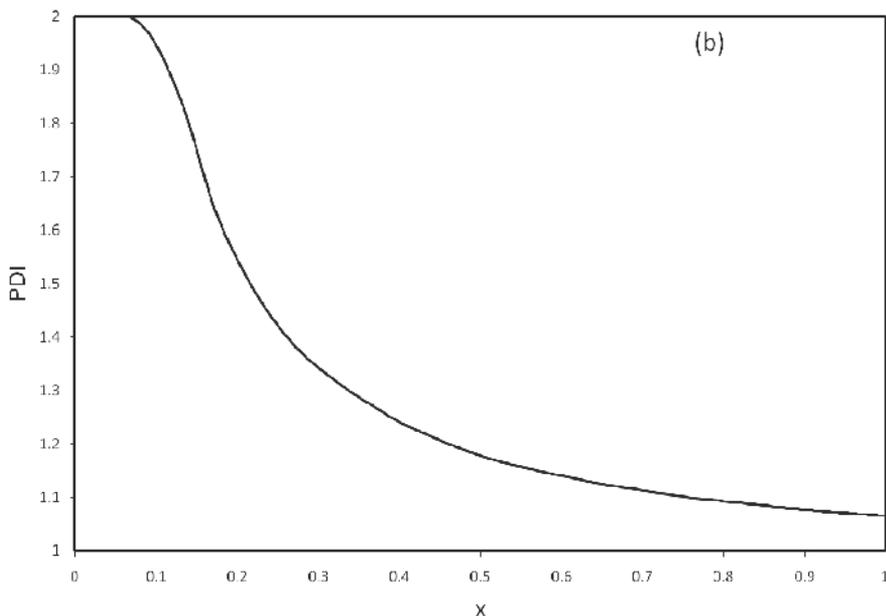


Fig. 4. Monte Carlo simulation results for the copolymerization of styrene and methyl methacrylate: PDI as a function of conversion. The initial comonomer molar fractions in the reactor were  $f_{0,St} = 0.75, f_{0,MMA} = 0.25$ . The rate constants and other parameters are tabulated in our journal paper (Al-harathi 2009a)

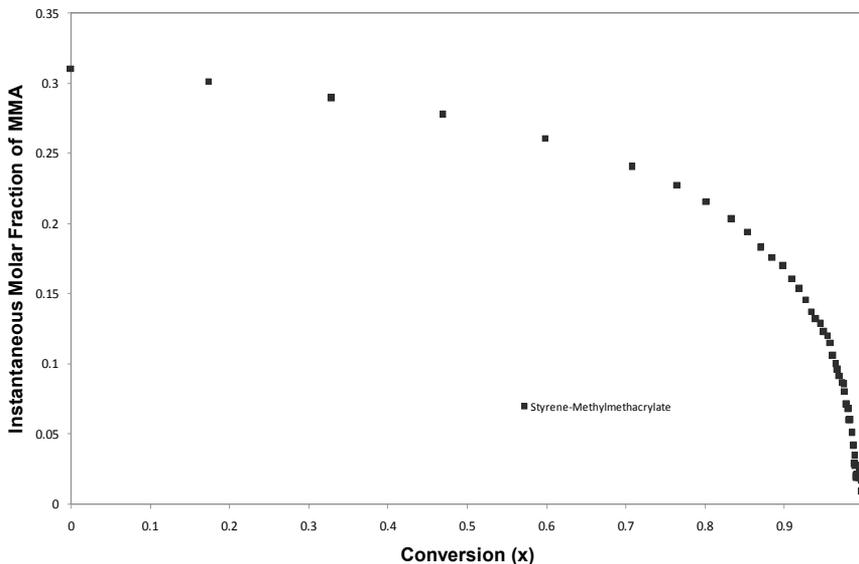


Fig. 5. Monte Carlo simulation results for the copolymerization of styrene- methyl methacrylate and acrylonitrile-methylmethacrylate copolymers : Instantaneous molar fraction as a function of conversion. The initial comonomer molar fractions in the reactor were  $f_{0,St} = 0.75, f_{0,MMA} = 0.25$ . The rate constants and other parameters are tabulated in our journal paper (Al-harathi 2009a)

Thus bimolecular termination reactions are programmed using the Monte Carlo approach. Other termination reactions can also be programmed with a similar methodology/ slightly modified approach as appropriate to the situation. Thus were some examples which were used to apply the generalized Monte Carlo approach for programming our physical system (ATRCp).

We would like to show some important results like PDI and instantaneous molar fraction. PDI is defined as the ratio of weight average molecular weight to the number average molecular weight. It is an important factor which governs the uniformity in the molecular weight of the polymer. Once we have simulated the termination reactions, most of the results can be predicted. Thus the correct place to show the result after explaining how to simulate the termination reactions. Figure 4 shows the simulated results for PDI obtained in ATRCp in a batch Reactor. We have shown that the PDI approaches a value of 1.1 in ATRCp which is coherent with the experimental data published in the field of ATRCp. PDI can be expressed mathematically in terms of means and standard deviation. We would like to encourage the students to search for the mathematical expression relating PDI in terms of means and standard deviations, hence we are not disclosing the simple mathematical expression. It also keeps the students interested and may also result in discovering other ways of incorporating the PDI. Figure 5 shows the instantaneous molar fraction of MMA in St-MMA produced in a batch reactor. The use of Mayo-Lewis equation can easily predict the instantaneous molar fraction.

#### 4. Conclusion

A thorough understanding of Monte Carlo procedure based on Gillespie's algorithm has been detailed. Specific MATLAB commands have been given for ease in programming. This would be of benefit to the readers as this is one of the first attempts giving out the actual procedure and programming technique for modeling free radical copolymerization. This procedure has been used by graduate students, researchers at King Fahd University of Petroleum and Minerals with great success. The authorial Al-harhi had developed the program and we have always wanted to educate others on this technique which has been the prime purpose of the article.

#### 5. References

- J. He, H. Zhang, J. Chen, Y. Yang, *Macromolecules* 1997, 30, 8010
- J. He, L. Li, Y. Yang, *Macromolecular Theory and Simulations* 2000, 9, 463
- H. Tobita, *Macromol Theory Simul.* 2003, 12, 32.
- H. Tobita *Macromolecules*, 1995, 28, 5119
- J. Lu, H. Zhang, Y. Yang, *Makro Chemie Theory Simul.* 1993, 2, 747.
- M. Al-Harhi, J. Soares, L. Simon, *Macromolecular Material Science* 2006, 291, 993
- M. Al-Harhi, J. Soares, L. Simon, *Macromolecular Reaction Engineering* 2007, 1, 95
- M. Al-Harhi, J. Masihullah, S.H. Abbasi, J. Soares *Macromolecular reaction engineering* 2009a in press
- M. Al-Harhi, J. Masihullah, S.H. Abbasi, J. Soares *Macromolecular Theory and Simulation* 2009b in press
- D. Gillespie, *J Phys Chem.* 1977, 81, 2340
- Editor Krzysztof Matyjaszewski - Book Controlled/Living Radical Polymerization From Synthesis to Materials (ACS Symposium Series)

# Monte Carlo Simulations in Solar Radio Astronomy

G. Thejappa<sup>1</sup> and R. J. MacDowall<sup>2</sup>

<sup>1</sup>*Department of Astronomy University of Maryland College Park MD 20742*

<sup>2</sup>*NASA/Goddard Space Flight Center Greenbelt MD 20771*

USA

## 1. Introduction

Propagation effects, especially, the refraction by the slowly varying ambient plasma, and the scattering by the random density fluctuations are known to distort the characteristics of the low frequency solar radio emissions. At kilometric radio wavelengths, they are probably responsible for the higher apparent source heights, larger source sizes, and widespread visibilities of type III radio bursts, and anomalous propagation time delays between signals arriving at widely separated spacecraft (Lecacheux et al , 1989; Steinberg et al , 1984; 1985; Thejappa et al , 2007). The scattering is also probably responsible for the low brightness temperatures and large equatorial diameters of the quiet sun (Aubier et al , 1971; Thejappa & Kundu , 1992; 1994; Thejappa and MacDowall , 2008a;b).

The regular refraction of radio waves in a spherically symmetric solar atmosphere has been investigated using the ray tracing methods (Bracewell & Preston , 1956; Jaeger & Westfold , 1950; Smerd , 1950; Thejappa and MacDowall , 2010). When the influence of plasma on wave propagation can be neglected, for example, in the case of interplanetary scintillations, the scattering and diffraction can be studied using parabolic equation methods (Bastian , 1994; Cairns , 1998; Lee & Jokipii, 1975; Rickett , 1977; Rytov et al , 1989). When the plasma can have a significant effect on the propagation, the geometric optics method is usually used to study the scattering of the solar radio emissions (Cairns , 1998; Thejappa et al , 2007). The scattering by multiple, independent and, random distribution of inhomogeneities is usually simulated using the statistical ray tracing techniques (Fokker , 1965; Hollweg , 1968; Riddle , 1974b; Steinberg et al , 1971; Steinberg , 1972). The geometric optics method treats the scattering as irregular refractions of rays, and introduces it as a random perturbation to the direction of the wave propagation vector. Such treatment is valid as long as the rms phase fluctuation  $\delta\phi = \frac{2\pi}{\lambda} \delta\mu \Delta S$  is not greater than a radian and the mean refractive index  $\mu$  is constant, so that the ray remains quasi-linear. Here, the  $\lambda$  is the wavelength,  $\mu$  is the refractive index, and  $\Delta S$  is the path length. The Monte Carlo methods are used to compute these small random perturbations in the directions of the rays due to scattering. For each scattering event, so that the scattering angles  $\langle \psi^2 \rangle = 2\sqrt{\pi} \int_{ray} \frac{\langle \delta\mu^2 \rangle}{\mu^2 l} dS$  are small and the rays remain quasilinear, the ray path is divided into linear steps of  $\Delta S$  chosen in such a way that the conditions  $\psi < 0.1$  radians, and  $\frac{\Delta\mu}{\mu} < 0.1$  are satisfied over each step. Here  $l$  is the scale length of the density fluctuations.

Fokker (1965) used the Monte Carlo technique to compute the sizes of the scatter images, intensity distributions, and directivities of type I solar radio bursts. By including the regular refraction into these statistical raytracing techniques, Steinberg et al (1971) and Riddle (1974b) studied the effects of scattering on the metric type III radio bursts. Steinberg (1972) used this technique to compute the directivity of type III bursts at 3 MHz, and compared with the observed center-limb histograms obtained by Fainberg & Stone (1970). The Monte Carlo methods are also used to study several other problems in solar radio astronomy (Aubier et al , 1971; Caroubalos et al , 1972; Hoang & Steinberg , 1977; Leblanc , 1973; Riddle , 1974a; Thejappa & Kundu , 1992; 1994; Thejappa et al , 2007; Thejappa and MacDowall , 2008a;b). For example, Aubier et al (1971) have shown that the scattering is probably responsible not only for lowering the brightness temperatures of the quiet sun radio emission at metric and decametric wavelengths, but also for raising their apparent source sizes. Thejappa & Kundu (1992) have shown that the scattering can lower the brightness temperatures of the quiet sun component to the observed values of  $\sim 10^5$  K, if the relative level of density fluctuations,  $\epsilon = \frac{\Delta N_e}{N_e}$  is at least of the order of  $\simeq 0.1$ . The main criticism of these studies is that they assume (1) idealized spherically symmetric density models for the solar atmosphere, and a Gaussian spectrum for the electron density fluctuations, and (2) arbitrary values for the relative level of density fluctuations,  $\epsilon$ , and their spatial scales,  $l$ . Several decades of in situ turbulence and interplanetary scintillation studies have yielded that the spatial power spectrum of density fluctuations is of a power-law type (Coles & Harmon, 1989; Coles et al , 1991). Similar extensive eclipse and coronagraph observations have shown that the spherically symmetric models for the electron density are highly idealized.

We have developed an efficient Monte-Carlo simulation technique, and applied it to study the directivity, visibility, time profiles, source sizes, and East-West asymmetries of low frequency type II and type III radio bursts (Thejappa et al , 2007; Thejappa and MacDowall , 2008b). We have also investigated the effects of refraction and scattering on the quiet sun radio emission (Thejappa and MacDowall , 2008a;b). Since, in the lower solar corona, the power is concentrated mainly in the flat part of the power spectrum with spectral index,  $\alpha = 3$ , we have derived an expression for the angular deflection suffered by a ray due to scattering by such density fluctuations in a slab of thickness,  $\Delta S$ . Using realistic models for the electron density, and density fluctuations, and observed values for  $\epsilon$ , and  $l$ , we statistically derive the emission characteristics of radio bursts, and quiet sun, and compare them with observations.

## 2. Model

### 2.1 Electron density

For the solar wind, we use the empirical formula derived by Bougeret et al (1984b)

$$N_e(r) = 6.14r^{-2.10} \text{ cm}^{-3}, \quad (1)$$

where  $r$  is the heliocentric distance in units of AU. For the quiet sun, we use the empirical formula derived by Guhathakurta et al (1996), based on Skylab data obtained during the declining phase of solar cycle 20 (1973-1976)

$$N_e(r, \theta_{mg}) = N_p(r) + [N_{cs}(r) - N_p(r)]e^{-\theta_{mg}^2/w^2(r)} \text{ cm}^{-3}, \quad (2)$$

where  $r$  is the heliocentric distance in units of  $R_\odot$ ,  $N_{cs}(r)$  and  $N_p(r)$  are the electron densities at the current sheet and the poles, respectively,  $w(r)$  is the half-angular width of the current

sheet,  $\theta_{mg}$  is the angular distance of a point from the current sheet in the heliomagnetic coordinate system (heliomagnetic latitude). The  $\theta_{mg}$  is given by

$$\theta_{mg} = \sin^{-1}[-\cos \theta \sin \alpha \sin(\phi - \phi_0) + \sin \theta \cos \alpha], \quad (3)$$

where  $\theta$  and  $\phi$  are the heliographic latitude and longitude, respectively,  $\alpha \simeq 15^\circ$  is the tilt angle of the dipole axis with respect to the rotation axis, and  $\phi_0 \simeq 0$  is the angle between the heliomagnetic and heliographic equators. The  $N_{cs}(r)$  and  $N_p(r)$  are defined as

$$N_e(r) = \sum_{i=1}^3 c_i r^{-d_i}, \quad (4)$$

where  $c_1, c_2$  and  $c_3$  are 1.07, 19.94, and 22.10 for the current sheet, and 0.14, 8.02, and 8.12 for the pole, respectively. These coefficients are in units of  $10^7$ . The coefficients  $d_1, d_2$  and  $d_3$  are 2.8, 8.45, and 16.87, respectively. The functional form of  $w(r)$  is

$$w(r) = \sum \gamma_i r^{-\delta_i}, \quad (5)$$

where  $\gamma_1 = 16.3^\circ$ ,  $\gamma_2 = 10^\circ$ ,  $\gamma_3 = 43.2^\circ$ ,  $\delta_1 = 0.5$ ,  $\delta_2 = 7.31$ , and  $\delta_3 = 7.52$ . We neglect the ambient magnetic field, because the electron cyclotron frequency  $f_{ce}$  is usually much less than the electron plasma frequency  $f_{pe}$ . We assume that the electron temperature  $T_e$  is  $1 \times 10^6$  K in the solar corona and  $1.5 \times 10^5$  K in the solar wind. We define the  $f_{pe}$ , the refractive index  $\mu$ , and the electron collision frequency  $\nu$  as

$$f_{pe}^2 = 80.6 \times 10^6 N_e, \quad (6)$$

$$\mu^2 = 1 - \frac{f_{pe}^2}{f^2}, \quad (7)$$

$$\nu = 4.36 N_e T_e^{-3/2} [17.72 + \ln(T_e^{3/2}/f)], \quad (8)$$

where  $f$  is the frequency in Hz.

## 2.2 Electron density fluctuations

The observations show that the spatial power spectrum of density fluctuations is of the power law type (Coles & Harmon, 1989; Coles et al, 1991)

$$P_n(q) = C_N^2 q^{-\alpha}; \quad q_o < q < q_i, \quad (9)$$

where  $q$  is the spatial wavenumber,  $\alpha$  is the spectral index, and  $l_o = 2\pi q_o$  and  $l_i = 2\pi q_i$  are the outer and inner scales of the density turbulence, respectively. Coles & Harmon (1989) have shown that for scales larger than a few times 100 km,  $\alpha$  is 11/3 (Kolmogorov spectrum), for intermediate scales (a few km  $\leq l \leq$  few times 100 km)  $\alpha$  changes from 11/3 to  $\sim 3$  (flat spectrum), and for the smallest scales of  $\sim 2$  km (inner or dissipative scales) the spectrum becomes quite steep with  $\alpha \simeq 4$ . The power is mainly concentrated in the flat part of the spectrum (Coles et al, 1991). By normalizing the spectrum to the variance of density fluctuations  $\langle \Delta N_e^2 \rangle$ , the expression for the structural constant  $C_N^2$  can be written as (Efimov et al, 2005)

$$C_N^2 = A(\alpha, q_o, q_i) \langle \Delta N_e^2 \rangle, \quad (10)$$

where

$$A(\alpha, q_o, q_i) = \begin{cases} \frac{(\alpha-3)\Gamma(\alpha/2)q_o^{\alpha-3}(2\pi)^{-3/2}}{\Gamma[(\alpha-1)/2]} & \text{for } 3 < \alpha < 4, \\ \frac{1}{4\pi \ln(\frac{2q_o}{q_i})} & \text{for } \alpha = 3. \end{cases}$$

For a Kolmogorov spectrum with  $\alpha = 11/3$ , this expression takes the form

$$C_N^2 = \frac{\epsilon^2 l_0^{-2/3} N_e^2}{6.6}, \quad (11)$$

where  $\epsilon = \frac{\Delta N}{N_e}$ . This expression agrees approximately with  $C_N^2 = \frac{\epsilon^2 l_0^{-2/3} N_e^2}{5.53}$ , derived by Spangler (2002). For the flat spectrum with  $\alpha = 3$ , this becomes

$$C_N^2 = \frac{\epsilon^2 N_e^2}{4\pi \ln(\frac{2l_i}{l_0})}. \quad (12)$$

For the solar wind,  $\alpha \sim 11/3$  agrees with that of Kolmogorov derived for fluid turbulence (Spangler & Sakurai, 1995; Spangler et al, 2002; Spangler, 2002; Tu & Marsch, 1994; Wohlmut et al, 2001; Woo et al, 1995). The inner scale  $l_i$  (which is also known as the dissipative scale) increases linearly with heliocentric distance as  $l_i = (\frac{R}{R_\odot})^{\pm 0.1}$  km at  $R \leq 100R_\odot$  and from 100 to 200 km,  $l_i \simeq 90 - 100$  km (Coles & Harmon, 1989; Manoharan et al, 1988). In this study, we assume that  $l_i \sim 100$  km, and for the outer scale  $l_0$ , we use the empirical formula derived by Wohlmut et al (2001) using the Galileo data between 7 to  $80 R_\odot$

$$l_0 = 19r^{0.82}. \quad (13)$$

Here we have rewritten the empirical relation of Wohlmut et al (2001) in units of AU. Based on *Helios* observations, Bavassano & Bruno (1995) deduced that most of the time  $\epsilon$  is 0.07 and is 0.1 for 14% of the time. In this study, we assume that  $\epsilon$  is 0.07 through out the solar wind (Cairns, 1998).

For the quiet sun studies, we consider the spatial scales, which range from 50 to 75 km. For  $l_i = 50$  km,  $l_0 = 75$  km and  $\alpha = 3$ , we obtain from (12)

$$C_N^2 = 0.28\epsilon^2 N_e^2. \quad (14)$$

The radio scattering observations indicate that the coronal and solar wind turbulence is highly anisotropic (see, for example, Armstrong et al (1990); Coles et al (2002); Grall et al (1997); Narayan et al (1989)). For example, Coles et al (2002) have given an empirical formula for the axial ratio, AR as

$$(AR - 1) \simeq \frac{160}{r^{3/2}}, \quad (15)$$

where  $r$  is in units of solar radii  $R_\odot$ . At heliocentric distances corresponding to meter and decameter wavelength radio emissions, we assume that the spatial scales along the magnetic field are 10 times larger than those perpendicular to the field. We also assume that  $\epsilon = 0.1$  throughout the corona.

### 3. Monte Carlo method

#### 3.1 Ray tracing

In the Cartesian coordinate system with origin at the center of the Sun, and the x-axis coinciding with the radial direction, the following set of 6 first-order differential equations describe the ray tracing (Haselgrove, 1963)

$$\frac{d\vec{R}}{d\tau} = \vec{T} \quad (16)$$

$$\frac{d\vec{T}}{d\tau} = D(\vec{R}) = \frac{1}{2} \frac{\partial \mu^2}{\partial \vec{R}}, \quad (17)$$

with

$$T_x^2 + T_y^2 + T_z^2 = \mu. \tag{18}$$

Here

$$\vec{R} \equiv \begin{pmatrix} x \\ y \\ z \end{pmatrix} \text{ and } \vec{T} \equiv \begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix}$$

are the position and direction vectors of the ray, respectively. The independent variable  $\tau$  is related to actual path length  $s$  as

$$d\tau = \frac{ds}{\mu}. \tag{19}$$

Using equations (6) and (7), we can write

$$D(\vec{R}) \equiv \frac{1}{2} \begin{pmatrix} \frac{\partial \mu^2}{\partial x} \\ \frac{\partial \mu^2}{\partial y} \\ \frac{\partial \mu^2}{\partial z} \end{pmatrix}$$

as

$$D(\vec{R}) = \frac{8.90 \times 10^{12}}{f^2} \frac{1}{r^4} N_e \vec{R}. \tag{20}$$

We use the Runge-Kutta algorithm to integrate the ray tracing equations (16) and (17), which can be written in vectorial form as (see for example Sharma et al (1982))

$$R_{n+1} = R_n + \Delta\tau [T_n + \frac{1}{6}(A + 2B)], \tag{21}$$

$$T_{n+1} = T_n + \frac{1}{6}(A + 4B + C), \tag{22}$$

$$A = \Delta\tau D(R_n), \tag{23}$$

$$B = \Delta\tau D(R_n + \frac{\Delta\tau}{2} T_n + \frac{1}{8} \Delta\tau A), \tag{24}$$

$$C = \Delta\tau D(R_n + \Delta\tau T_n + \frac{1}{2} \Delta\tau B), \tag{25}$$

$$D(R) = \frac{1}{2} \frac{\partial \mu^2}{\partial R}. \tag{26}$$

We compute the optical depth  $\tau$  and the transit time  $\Delta t$ (s) at each step  $\Delta S$  on the ray path as

$$\tau_{i+1} = \tau_i + K \frac{f_{pe}^2}{f^2} \frac{v \Delta S}{\mu_i} \tag{27}$$

$$\Delta t_{i+1} = \Delta t_i + K \frac{\Delta S}{\mu_i}, \tag{28}$$

where  $K$  is 500, (1 A. U divided by velocity of light) in the solar wind, and 2.32 (solar radius divided by velocity of light) in the solar corona. Using this algorithm, one can trace the rays through any medium, i.e., starting from a known point  $(\vec{R}_0, \vec{T}_0)$ , one can generate successively  $(\vec{R}_1, \vec{T}_1), (\vec{R}_2, \vec{T}_2).....(\vec{R}_n, \vec{T}_n)$ .

### 3.2 Scattering

The components of  $\vec{T}$  (eqn. 22) are the direction cosines of the ray after it has suffered a regular refraction in a layer of thickness  $\Delta S$ . The scattering by random density fluctuations in this layer, i.e., the random perturbation vector  $\langle \vec{p} \rangle$  is added to  $\vec{T}$  at each step. The components of  $\langle \vec{p} \rangle$  are computed using the mean-square angular deviation  $\langle \Psi^2 \rangle$  suffered by the ray due to scattering in the layer of  $\Delta S$ . Chadrashkar (1952) and Hollweg (1968) have derived the expression

$$\langle \Psi^2 \rangle = b(f) \Delta S, \quad (29)$$

where  $b(f)$  is the mean square deviation per unit length. For the power spectrum  $P_n(q) = C_N^2 q^{-\alpha}$ , the expression for the mean-square angular deviation is (Cairns, 1998; Thejappa et al, 2007)

$$\langle \Psi^2 \rangle = \frac{r_e^2 \lambda^4}{\pi \mu^2} \Delta S \frac{C_N^2}{4 - \alpha} (q_i^{4-\alpha} - q_o^{4-\alpha}). \quad (30)$$

For  $q_o \ll q_i$ , this expression can be simplified as

$$\langle \Psi^2 \rangle = \frac{r_e^2 \lambda^4}{\pi \mu^2} \Delta S C_N^2 \frac{q_i^{4-\alpha}}{4 - \alpha}. \quad (31)$$

For  $\alpha = 11/3$  (spectral index),  $r_e = \frac{e^2}{mc^2}$  (the classical radius of the electron),  $\lambda = \frac{c}{f}$  (the wavelength of the wave in free space),  $C_N^2 = \frac{\epsilon^2 l_a^{-2/3} N_e^2}{5.53}$ , (structural coefficient), and  $f_{pe}^2 = \frac{\epsilon^2 N_e}{\pi m_e}$  (the electron plasma frequency), we obtain from equations (29) and (31)

$$b(f) = \pi \frac{f_{pe}^4}{f^4} \frac{\epsilon^2}{\mu^4 l_i^{1/3} l_o^{2/3}}. \quad (32)$$

For the flat spectrum with  $\alpha = 3$ , and for  $Q_i = \frac{2\pi}{L_i}$  and  $Q_o = \frac{2\pi}{L_o}$  (corresponding to the lower and upper limits of the range of the considered spatial scales in the place of  $q_i$  and  $q_o$ ) we obtain from equation (30)

$$b(f) = \pi \frac{(1 - L_i/L_o) f_{pe}^4}{2 \ln(2L_i/L_o) f^4} \frac{\epsilon^2}{\mu^4 L_i}, \quad (33)$$

where, we have used the expression  $C_N^2 = \frac{\epsilon^2 N_e^2}{4\pi \ln(2L_i/L_o)}$ . For  $L_i = 50$  km, and  $L_o = 75$  km, we obtain from equation (33)

$$b(f) \sim 0.6\pi \frac{f_{pe}^4}{f^4} \frac{\epsilon^2}{\mu^4 L_i}. \quad (34)$$

For Gaussian fluctuations, the expression for  $b(f)$  (Lacombe et al, 1988)

$$b(f) = \frac{\sqrt{\pi} f_{pe}^4}{\mu^4} \frac{\epsilon^2}{f^4 h}, \quad (35)$$

coincides with the expression (32) derived for the Kolmogorov spectra for an effective scale height  $h = l_i^{1/3} l_o^{2/3}$ . Similarly, the expression (34) derived for  $\alpha = 3$  coincides with (35) derived for the Gaussian spectrum. The components of  $\langle \vec{p} \rangle$  are chosen from a Gaussian distribution of random numbers with a zero mean and a standard deviation of

$$\sigma = \mu \sqrt{b \Delta S}. \quad (36)$$

For isotropic fluctuations, three independent Gaussian distributed random deviations of the direction cosines with the same standard deviation (36) are calculated. However, for anisotropic fluctuations with longitudinal scales much larger than the transverse scales,  $\sigma$  changes accordingly, with  $\sigma_{\parallel} < \sigma_{\perp}$ , since  $\sigma$  is inversely proportional to the square root of the spatial scale of the density fluctuations.

#### 4. Type III radio bursts

By assuming that both fundamental and harmonic emissions are emitted by isotropic point sources, we can assign the initial directions of the rays so that they end on a sphere of radius equal to the local  $\mu_0$ . Then, the probability  $p(\theta_0, \phi_0)$ , that a point belongs to an element of a spherical surface  $\sin \theta_0 d\theta_0 d\phi_0$  is  $\frac{\sin \theta_0 d\theta_0 d\phi_0}{4\pi}$ , where  $0 \leq \theta_0 \leq \pi$  and  $0 \leq \phi_0 \leq 2\pi$ . By writing  $p(\theta_0, \phi_0) = p_1(\theta_0)p_2(\phi_0)$ , we obtain  $p_1(\theta_0) = \frac{\sin \theta_0}{2}$  and  $p_2(\phi_0) = \frac{1}{2\pi}$ , and from the integrals  $\int_0^{\phi_0} p_2(\phi_0) d\phi_0 = \frac{\phi_0}{2\pi} = \xi_1$ , and  $\int_0^{\theta_0} p_1(\theta_0) d\theta_0 = \frac{1}{2} \int_0^{\theta_0} \sin \theta_0 d\theta_0 = \frac{\cos \theta_0}{2} + \frac{1}{2} = \xi_2$ , we obtain the azimuthal and elevation angles of the initial ray directions as

$$\phi_0 = 2\pi\xi_1 \quad (37)$$

$$\cos \theta_0 = 2\xi_2 - 1, \quad (38)$$

where  $\xi_1$  and  $\xi_2$  are the random variables distributed uniformly between 0 and 1. Thus, the optical direction cosines of initial rays can be written as

$$T_{x0} = \mu_0 \sin \theta_0 \sin \phi_0 \quad (39)$$

$$T_{y0} = \mu_0 \sin \theta_0 \cos \phi_0 \quad (40)$$

$$T_{z0} = \mu_0 \cos \theta_0, \quad (41)$$

which imply that  $-\mu_0 \leq (T_{x0}, T_{y0}, T_{z0}) \leq \mu_0$ . By assigning these random directions for the directions of initial rays, we eliminate the possibility for any two rays to have the same initial direction.

We have developed a computer program based on the algorithm presented in the previous section. We have used it to trace the rays statistically by assuming that both fundamental and harmonic emissions are emitted at 120 kHz, corresponding to altitudes of 0.2097 AU ( $f_{pe} \sim 115$ ) and 0.3895 AU ( $f_{pe} = 60$  kHz), respectively. We have considered two cases (1)  $\epsilon = 0$  and (2)  $\epsilon = 0.07$ . For each case, we launch 1000 randomly directed rays and trace them until they cross the sphere of 1 AU radius. For  $\epsilon = 0$ , when only the regular refraction is considered, we take  $\Delta S = 0.002$  AU, and for  $\epsilon = 0.07$ , when both regular refraction and scattering are dominant we take  $\Delta S = 10l$ , where  $l = l_i^{1/3} l_o^{2/3}$ . At the exit point, we record the components of  $(\vec{R})$  and  $(\vec{T})$ , total optical depth  $\tau$  and time delay  $\Delta t$  (calculated using equations 27 and 28) in a separate file. These recorded values will be used to calculate the directivities, time profiles, and sizes and heights of the apparent sources of both fundamental and harmonic emissions. In Fig. 1, we present the typical trajectories of the traced rays. Here, the first and second columns correspond to the fundamental and harmonic emissions, whereas the first and second rows correspond to the cases,  $\epsilon = 0$ , and  $\epsilon = 0.07$ , respectively. We have embedded the distributions of the ray trajectories inside the transparent spheres of 1 AU radius to have a better representation of the Sun, the rays emanating from the source, and the observer. The distributions of the rays in the first row show that the regular refraction focuses the fundamental into a narrower cone than that of the harmonic, and the distributions in the second row show that the scattering ( $\epsilon = 0.07$ ) destroys the refractive focusing.

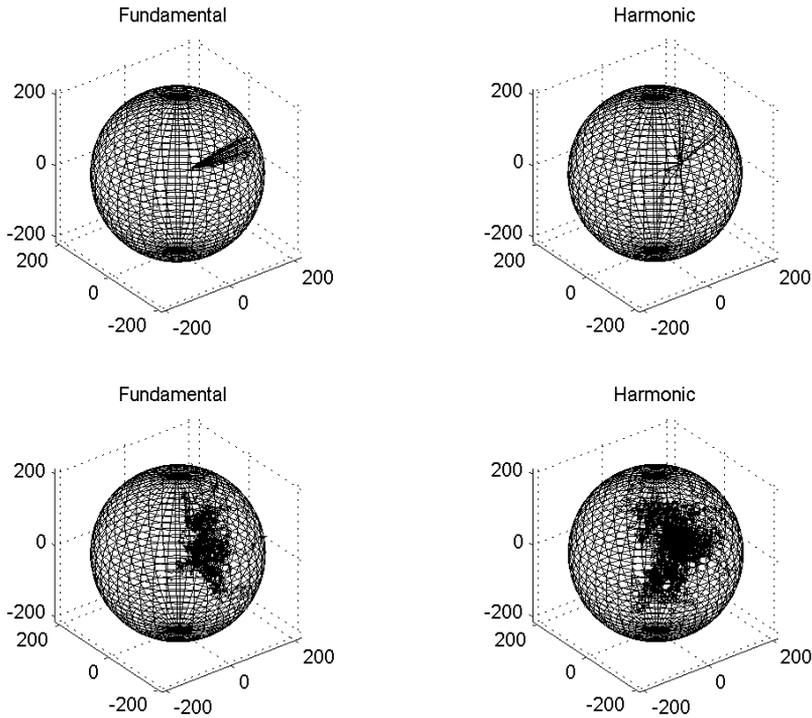


Fig. 1. Typical distributions of the traced rays of the fundamental (first column) and harmonic (second column) emissions. The first row corresponds to the case of only regular refraction, whereas the second row corresponds to the regular refraction as well as scattering.

#### 4.1 Directivity

Each ray is characterized by an angle  $\beta$  subtended at the center by the traced portion of the ray from the source to 1 AU. Since in the spherically symmetric case, the axis through the source and the center of the Sun is also the axis of the cylindrical symmetry, the angle  $\beta$  determines the distance of the apparent source from the center of the solar disk. In heliographic degrees,  $\beta$  is defined as

$$\beta = \frac{\cos^{-1}(\vec{R} \cdot \vec{x})}{|\vec{R}|}, \quad (42)$$

where the x-axis is the axis of symmetry, and  $\vec{R}$  is the position vector at the exit point. For a source on the solar equator,  $\beta$  is the longitude measured from the central meridian. The directivity is defined as the ratio of the power received (i.e., number of rays) in a range of angles from  $\beta$  to  $\beta + d\beta$  from the source embedded in a scattering and refracting medium, to the power received from the same source at the same position emitting the same total power isotropically in a vacuum. This can be expressed as a ratio of the total number of rays in a group of angles centered around  $\beta$  each weighted by  $e^{-\tau}$ , to the total number of rays that would fall in the same group of  $\beta$  from an isotropic source when the ray paths are unaffected by either scattering or refraction, i.e.,

$$D(\beta) = \frac{4\pi \sum n_{\beta} e^{-\tau_{\beta}}}{\Delta \Omega N_T}, \quad (43)$$

where  $n_\beta$  is the number of rays escaping in the angles from  $\beta$  to  $\beta + d\beta$ , and  $N_T$  is the total number of rays. The total optical depth  $\tau_\beta$  along each ray is computed by summing the optical depths along all the steps taken by the ray from the source to its exit as given by equation (27). The attenuation coefficient  $e^{-\tau}$  represents the losses suffered by the ray due to increased path lengths caused by scattering. The solid angle  $\Delta\Omega$  spanned by grid separation in the  $\beta$  direction around the annular ring is defined as

$$\Delta\Omega = 4\pi \sin[(i_\beta + 0.5)\Delta\beta] \sin(\Delta\beta/2), \tag{44}$$

where  $\Delta\beta$  and  $i_\beta$  are the angular width and index of the group, respectively. We have computed the directivities for  $\epsilon = 0$  as well as for  $\epsilon = 0.07$  by counting the number of harmonic and fundamental rays in groups of 5 and 1 degree intervals, respectively, and normalized them by dividing each of them by the largest in each case. In Fig. 2, we present these normalized directivities, where the first row clearly shows that in a smoothly varying plasma ( $\epsilon = 0$ ), the refraction focuses the fundamental and harmonic emissions into cones of  $\sim 18^\circ$  and  $\sim 80^\circ$  angular widths, respectively. The refractive focusing can be understood in

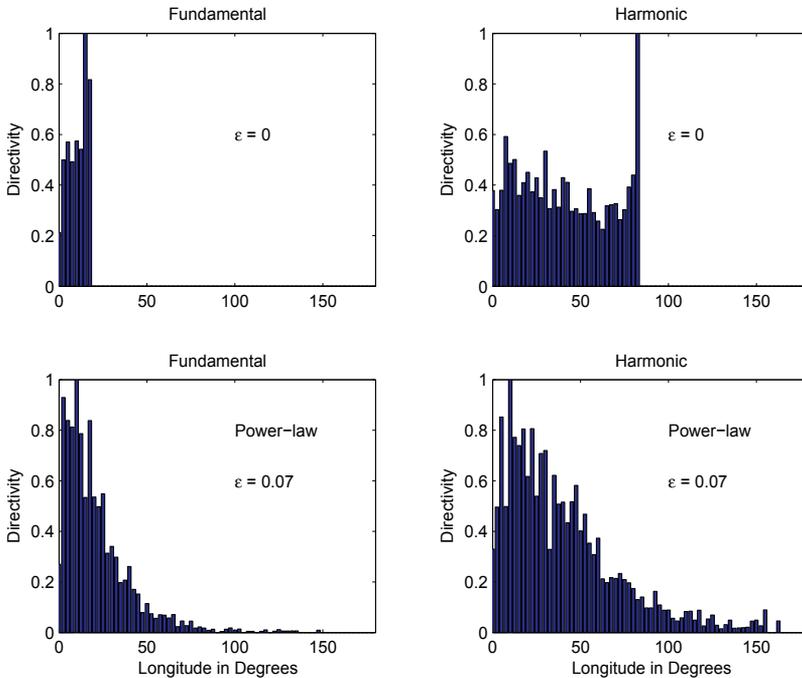


Fig. 2. The left and right columns show the directivities of the fundamental and harmonic emissions, respectively. Here  $\epsilon$  is equal to the level of relative density fluctuations  $\frac{\Delta N_e}{N_e}$

terms of the law of refraction in a plane-layered plasma

$$\mu(z) \sin \phi = \mu(z_0) \sin \phi_0, \tag{45}$$

where  $\phi$  is the angle between the ray and the gradient of the refractive index. At the exit point  $z = z_0$ , the refractive index  $\mu(z_0)$  is  $\sim 1$  and the angle  $\phi$  is  $\phi_0$ , and at the point of reflection  $z = z_{ref}$  the angle  $\phi$  is  $\pi/2$  and (from equation (45))

$$\mu(z_{ref}) = \sin \phi_0. \tag{46}$$

For an isotropic point source, the refraction bends the rays into a cone of angular width  $\phi_0$ . The apex of the ray leaving the plasma at the maximum angle  $\phi_0 = \phi_{max}$  coincides with the source location  $z = z_s$ . The angular width of the limiting cone can be obtained from equation (46) as

$$\phi_{max} = \sin^{-1} \mu(z_s) = \sec^{-1}(f/f_{pe}). \quad (47)$$

For example,  $\phi_{max}$  is  $5^\circ$ ,  $16.6^\circ$ , and  $60^\circ$ , for  $f \sim 1.004f_{pe}$ ,  $f \sim 1.0435f_{pe}$ , and  $f = 2f_{pe}$ , respectively. The computed limiting cones (see first row of Fig. 2) are slightly broader than those estimated using equation (47), because in the computations spherical symmetric model is used, whereas the  $\phi_{max}$  in equation (47) is derived using the plane parallel approximation. The intense "shoulders" at the edges of the limiting cones seen in the directivity diagrams (first row of Fig. 2) at  $\simeq 18^\circ$  and  $\simeq 80^\circ$  longitudes probably are due to ingoing rays from the source (Steinberg et al , 1971; Steinberg , 1972).

We use the histograms presented in row 1 to compute the directivity factors as the ratios of intensities at two different longitudes. For  $\epsilon = 0$ , the ratio of intensities at  $5^\circ$  and  $15^\circ$  is 0.5 for the fundamental, and the ratio of intensities at  $5^\circ$  and  $80^\circ$  longitudes is 0.75 for the harmonic. For  $\epsilon = 0$ , the fundamental emission is very intense and directive for an observer located within its limiting cone in comparison with that of the harmonic, by indicating that during low level of density fluctuations, the positive identification of the mode of emission of the observed type III or the type II bursts with the fundamental, or a mixture of strong fundamental and weak harmonic emissions is a good indication of an imminent arrival of the flare accelerated electrons or the CME driven shock accelerated electrons at the spacecraft.

#### 4.2 Time profiles

The arrival times of scattered rays ultimately received in a given direction are different for different rays. If the source radiates a very short pulse, the intensity recorded as a function of time is the "transient" response of the ambient medium. Thus, the observed time profile can be understood as the convolution of the time profile at the source and the time response at 1 AU of an impulsive burst at the source. The ambient medium refracts and scatters the ray from the time it is launched until it exits the medium. The time taken by the ray during each step  $\Delta t_i$  is  $\frac{S_i}{c\mu_i}$ , where  $c$  is the velocity of light,  $S_i$  is the path length traveled by the ray during the  $i$ -th step, and  $\mu_i$  is the refractive index; the  $c\mu_i$  is the group speed of the ray. The time taken by the ray to travel from the source to the point of exit from the medium (arrival time) is the sum of all the time steps  $\Delta t_i$  as given in equation (28). The histogram of these arrival times at 1 AU gives the time profile of the isotropic point source. In the first column of Fig. 3, we present the time profiles of the fundamental constructed using the arrival times of the rays gathered in the longitude range from 0 to  $30^\circ$ . These show that the total durations of the unscattered and scattered emissions are 50 (top panel), and  $\sim 3000$  seconds (bottom panel), respectively. The time profile of the scattered emission, which is characterized by a rapid rise followed by an exponential decay resembles the time profile of an idealized type III radio burst. The exponential decay can be written as  $\exp(-t\nu_{120})$ , where  $\nu_{120}$  is the effective collision rate due to scattering of 120 kHz fundamental. From Fig. 3, we can estimate that  $1/\nu_{120} \sim 2000$  sec. In the second column of Fig. 3, we present the histogram of the arrival times of the harmonic rays gathered in the longitude range from 0 to  $90^\circ$ . These time profiles show that the durations of the unscattered and scattered emissions are  $\sim 350$  and  $\sim 3000$  seconds, respectively. The time profile of the scattered harmonic emission also appears like that of an idealized type III burst time profile. In this case the law for the exponential decay can be written as  $\exp(-2t\nu_{60})$ , where  $\nu_{60}$  is the effective collision rate due to scattering of 120

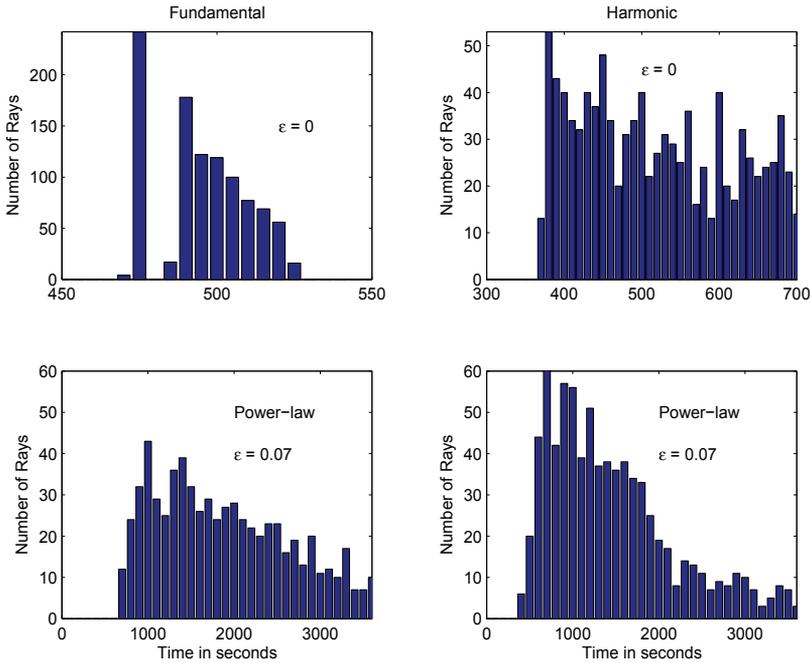


Fig. 3. The computed time profiles of the fundamental (left column) and harmonic emissions (right column) at 120 kHz for various cases. The  $\epsilon$  is equal to the level of relative density fluctuations  $\frac{\Delta N_e}{N_e}$

kHz harmonic excited at 60 kHz critical layer. Again, from Fig. 3, we can estimate that  $1/v_{60}$  is  $\sim 1500$  sec. The computed durations of  $\sim 3000$  s for both fundamental as well as harmonic emissions are comparable to the observed durations of type III bursts at these low frequencies (Steinberg et al , 1984).

The computed time profiles also show that the fundamental (F) and harmonic (H) emissions arrive at the spacecraft at different times, i.e., at a given frequency a time profile may contain two peaks corresponding to F and H emissions. However, the excitation of the harmonic emission is delayed with respect to the fundamental. This delay  $\Delta t$  is equal to the time taken by the beam to travel from the plasma level ( $\sim f_{pe}$ ) to  $\frac{f_{pe}}{2}$  level

$$\Delta t \simeq \frac{r(f_{pe}) - r(f_{pe}/2)}{v_b}, \quad (48)$$

where  $v_b$  is the beam speed. We use equations (1), and (6), and write the heliocentric distance as a function of the electron plasma frequency as

$$r(f_{pe}) = 19.2 f_{pe}^{-0.952} \text{ AU}. \quad (49)$$

This can be used to calculate  $\Delta t$  for a given value of  $v_b$ . For example, a beam traveling at speeds of  $v_b = \alpha c$  takes  $\frac{89.9}{\alpha}$  seconds to travel from the fundamental layer at 0.2097 AU (corresponding to  $f_{pe} = 115$  kHz) to the harmonic layer at 0.3895 AU (corresponding to  $f_{pe} = 60$  kHz). Thus, for  $\alpha$  equals to 0.1 and 0.5, the time delay is  $\sim 15$  and  $\sim 3$  minutes, respectively. These values are consistent with the observed time delays of the fundamental-harmonic pairs

at high frequencies (Caroubalos & Steinberg , 1974), as well as with those of interplanetary type III radio bursts. For typical beam velocities, the harmonic peak always occurs later than the fundamental. However, due to superposition only a single peak may appear in the time profile. Such a peak may correspond to the fundamental at the low longitudes, and to the harmonic at high longitudes. The occurrence of the fundamental emission followed by the harmonic is consistent with observations (Dulk et al , 1984; Kellogg , 1980), which implies that the time profile of a type III burst may contain a single emission peak, consisting of a mixture of the fundamental (rise part) and harmonic (peak and decay part) components. Thus, the computed time profiles and time delays between the fundamental and harmonics explain the observations that the type III bursts in the solar wind often have both fundamental and harmonic components, that in some bursts and some frequencies, the fundamental component is dominant, in others only the harmonic component is present, and in many there are two components but the components overlap considerably (Dulk et al , 1984; Kellogg , 1980; Reiner & Stone , 1988; 1989; Reiner et al , 1992; Thejappa et al , 1993).

### 4.3 Source size and displacement

We project the exit points ( $\vec{R}$ ) of the rays in the angular range  $\beta$  and  $\beta + d\beta$  onto a plane passing through the source S and perpendicular to the exit direction,  $\vec{T}$ . The distributions of the projected points determine the sizes and displacements of the apparent sources. The equation of the plane through the source with radius vector  $R_s$  and normal to  $\vec{T}$  can be written as

$$T_1x + T_2y + T_3z = D, \quad (50)$$

where  $D = T_1x_1 + T_2y_1 + T_3z_1$ , and  $(x_1, y_1, z_1)$  and  $(T_1, T_2, T_3)$  are the components of the vectors  $\vec{R}_s$  and  $\vec{T}$ , respectively. The projection of the exit point  $(x_2, y_2, z_2)$  on this plane can be obtained from equation

$$\frac{x - x_2}{T_1} = \frac{y - y_2}{T_2} = \frac{z - z_2}{T_3} = p, \quad (51)$$

as  $(pT_1 + x_2, pT_2 + y_2, pT_3 + z_2)$ , where  $p$  is a parameter. By substituting these coordinates in the equation of the plane (50), we obtain

$$\vec{T} \cdot (p\vec{T} + \vec{R}) = D = \vec{T} \cdot \vec{R}_s \quad (52)$$

$$p = \frac{\vec{T} \cdot (\vec{R}_s - \vec{R})}{\vec{T} \cdot \vec{T}}. \quad (53)$$

In the first and second columns of Fig. 4, we present the distributions of projected points of the fundamental and harmonic emissions scattered into the longitude range of 0 to 30°. These distributions represent the sizes of the apparent sources. The altitudes of these apparent sources can be computed as the heliocentric distances of the centroids of these distributions. For example, the 120 kHz fundamental source located at 0.2097 AU (corresponding to  $f_{pe} = 115$  kHz) is displaced inward to the radial distance of 0.2033 AU (the critical layer corresponding to  $\sim 120$  kHz) in the absence of density fluctuations, i.e., when  $\epsilon = 0$ . On the other hand, the fundamental source is displaced radially outward to a distance of 0.5950 AU (critical layer corresponding to 38.8 kHz) due to scattering by density fluctuations with  $\epsilon = 0.07$ . Thus, the apparent source of the fundamental lies at a radial distance corresponding to  $\sim \frac{f}{3}$  layer. This altitude of the centroid of the apparent fundamental source agrees very well with the observed heights of  $f/2$  and  $f/5$  layers for type III radio bursts (Steinberg et al , 1985). In the harmonic case, the computed centroid of the apparent source shows that it is

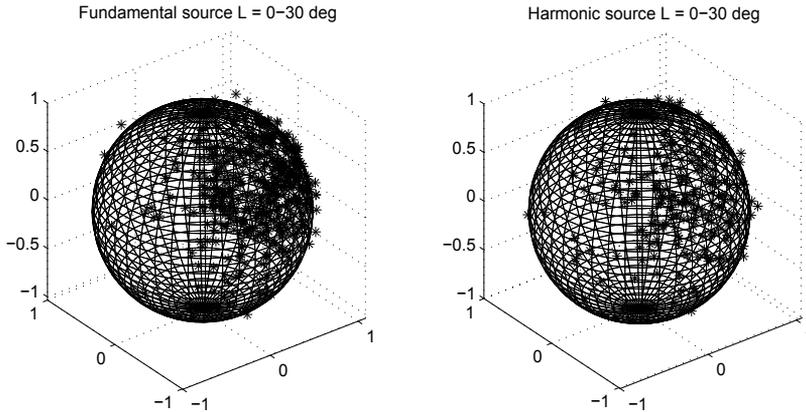


Fig. 4. The projected images of the scattered fundamental and harmonic sources.

		Fundamental			Harmonic		
Angles	$\epsilon = \frac{\Delta N_e}{N_e}$	$S_{\parallel}$	$S_{\perp}$	Location	$S_{\parallel}$	$S_{\perp}$	Location
0-30	$\epsilon = 0$	0.2429	1.1233	0.2033	6.5	1.96	0.3152
0-30	$\epsilon = 0.07$ (Power-law)	24	27	0.5950	36	38	0.3993
0-60	0	—	—	—	6.22	2.13	0.3211
0-60	power-law	—	—	—	36	38	0.3993

Table 1. The radial distances (location) in units of AU, the radial  $S_{\parallel}$  and transverse  $S_{\perp}$  sizes (in degrees) of the refracted and scattered fundamental and harmonic sources at 120 kHz. The first column shows the range of longitudes used for these estimates

displaced inwards from its initial location of 0.3895 AU corresponding to 60 kHz plasma level to 0.3152- 0.3329 AU in the absence of density fluctuations ( $\epsilon = 0$ ). These radial distances correspond to  $f_{pe} \sim 0.6f$ . When the scattering is included in the computations with  $\epsilon = 0.07$ , the height of the apparent harmonic source is displaced to a radial distance of 0.3993 AU, corresponding to the  $\sim \frac{f}{2}$  critical level. The sizes of the apparent sources are computed as the half-power widths of these distributions. When only refraction is considered with  $\epsilon = 0$ , the size of the apparent fundamental source is  $0.2429^{\circ}$ , and  $1.1233^{\circ}$  parallel and perpendicular to the radial direction. On the other hand, when the scattering is included with  $\epsilon = 0.07$ , the size of the apparent source is increased to  $24^{\circ}$  and  $27^{\circ}$  parallel and perpendicular to the radial direction, respectively. To estimate the sizes of the harmonic sources, we have considered two ranges of longitudes, namely, 0 to  $30^{\circ}$  and  $30^{\circ}$  to  $60^{\circ}$ . When  $\epsilon = 0$ , the size of the apparent harmonic source is  $\sim 6.5^{\circ}$ , and  $\sim 2^{\circ}$  along and across the radial directions in both longitude ranges. When the scattering is included, the sizes in both cases are increased to  $36^{\circ}$  and  $38^{\circ}$  along and across the radial direction, respectively. In Table 1, we present the computed sizes and the heliocentric distances of the apparent sources. These computed sizes and heights agree very well with the observations of Steinberg et al (1984), who after analyzing a large data set reported that (1) the angular sizes of type III sources vary from  $\sim 5^{\circ}$  at 1000 kHz to  $\sim 50^{\circ}$  at 100 kHz, and  $\sim 60$  percent of all 100 kHz angular sizes were between  $40^{\circ}$  and  $60^{\circ}$ , and (2) the heliocentric distances of type III source centroids at a given frequency  $f$  range from the distance where  $f_{pe} = f/2$  to that where  $f_{pe} = f/5$ .

## 5. Comparison with observations

In Fig. 5, we present an example of a multi spacecraft detection of a type II and a couple of type III radio bursts by the Unified Radio and Plasma Wave (URAP) experiment on Ulysses (Stone et al , 1992) and the Waves investigation on Wind (Bougeret et al , 1995). Ulysses is in a highly elliptical orbit out of the ecliptic plane with aphelion (perihelion) at  $\sim 5.4$  AU ( $\sim 1.3$  AU), the trajectory of the Wind takes it from near Earth orbits to the Lagrange point (L1), about  $230 R_E$  upstream of Earth. The Ulysses data presented in the top panel show an intense type III burst after 12:00 on 1997/11/6 and several other weaker type III bursts. The data early on 1997/11/6 are corrupted by a poor telemetry link. The type II emission is the weaker activity (see color bar scale) starting at 18:00 and continuing to 12:00 on the next day while drifting from 200 to 100 kHz. The flare site related to these events was at S18 and W63 according to the Solar Geophysical Data. The bottom panel shows similar data from the Wind spacecraft, where the same type II and type III bursts are seen; they are detected slightly earlier because Wind is closer to the Sun (at 1 AU near the Earth) than Ulysses (at 5.3 AU). The additional emission features in the Wind data are Auroral kilometric Emission (AKR -a terrestrial radio emission appearing as short duration, broad-band feature throughout the plot) and the electron thermal noise (the horizontal feature seen in the bottom half of the panel). In the middle panel, single frequency data from the two spacecraft at  $\sim 120$  kHz is plotted. The signal levels are not the same at the two spacecraft because (1) the radio bursts are directive, and (2) the distances of the sources are different for different spacecraft. However, the time profiles of type III as well as type II bursts observed at Ulysses are very similar to those observed at WIND. Note that the Ulysses data plotted in the middle panel have been shifted in time by about 35 minutes to correct for the longer propagation distance to Ulysses. (ULYSSES: Heliographic latitude =  $2^\circ$ , Heliographic longitude =  $53.9^\circ$ , Range to Sun = 5.3 AU; EARTH/WIND: Heliographic latitude =  $3.8^\circ$ , Heliographic longitude =  $301.2^\circ$ , Range to Sun = 1.0 AU).

We compute the distributions of rays emitted by the fundamental and harmonic sources located at (S18, W63) at altitudes of 0.2050 and 0.3895 AU and examine whether they are visible to Ulysses and Wind spacecraft, We trace the rays corresponding to both F and H components until they reach the distances comparable to those of Ulysses for (1)  $\epsilon = 0$  as well as (2)  $\epsilon = 0.07$ . In Fig. 6, we show the typical distributions of these traced rays, where we also show the locations of the Ulysses and Wind spacecraft. It is clear from these distributions of traced rays that (1) when  $\epsilon = 0$  the fundamental is highly beamed and visible only to Ulysses spacecraft, (2) when  $\epsilon = 0.07$  the scattering causes the fundamental to be visible to Ulysses as well as Wind spacecraft by destroying the limiting cone, and (3) the harmonic emission is visible to both Ulysses and Wind spacecraft for  $\epsilon = 0$  as well as for  $\epsilon = 0.07$ . Thus, except for the refracted fundamental, the rest of the emissions, namely scattered fundamental, unscattered harmonic as well as scattered harmonic can be visible to both Ulysses and Wind spacecraft. This indicates that the visibility of radio bursts critically depends on the coordinates of their sources.

## 6. Quiet sun component

The brightness temperature of the thermal emission from the quiet Sun is computed as

$$T_b = T_e(1 - e^{-\tau}), \quad (54)$$

where

$$\tau = \int_{s_1}^{s_2} \zeta ds \quad (55)$$

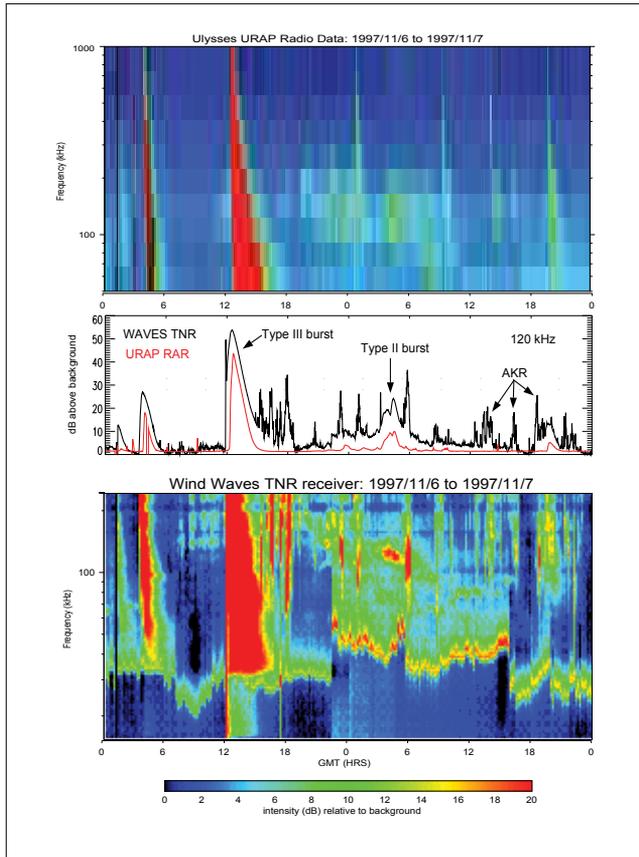


Fig. 5. Top: In the top panel, Ulysses URAP data show an intense type III burst after 12:00 on 1997/11/6, as well as several other weaker type IIIs. The type II emission is the weaker emission band (see color bar scale) from 18:00 to 12:00 on the next day and drifting from 200 to 100 kHz. The bottom panel shows similar data from the Wind Waves instrument. In the middle panel, single frequency data from the two spacecraft at approximately 120 kHz is plotted. Because of differing distances from the sources to the spacecraft as well as the effects of directivity, the signal levels seen for a given emission is different at the two spacecraft. Bottom: The two spacecraft are separated by more than  $100^\circ$  in heliographic longitude, providing an ideal angular separation for studying these events.

is the optical depth, and  $s_1$  and  $s_2$  are the heliocentric distances of the source and the observer, respectively. The absorption coefficient per centimeter of path length,  $\zeta$ , is defined as

$$\zeta = \frac{f_{pe}^2 v}{f^2 \mu c}. \quad (56)$$

The brightness temperature at some point on the solar disk is determined by using equation (54), where the optical depth  $\tau$  is calculated by tracing the rays (initially launched toward that point). The rays are traced from a distance of  $2.5R_\odot$  toward the Sun until the optical depth reaches a large value of  $\sim 10$ , or the ray is traveled at least  $5R_\odot$ . The rays are launched only in the equatorial plane, where the x-axis is directed toward the observer. Here, the positive

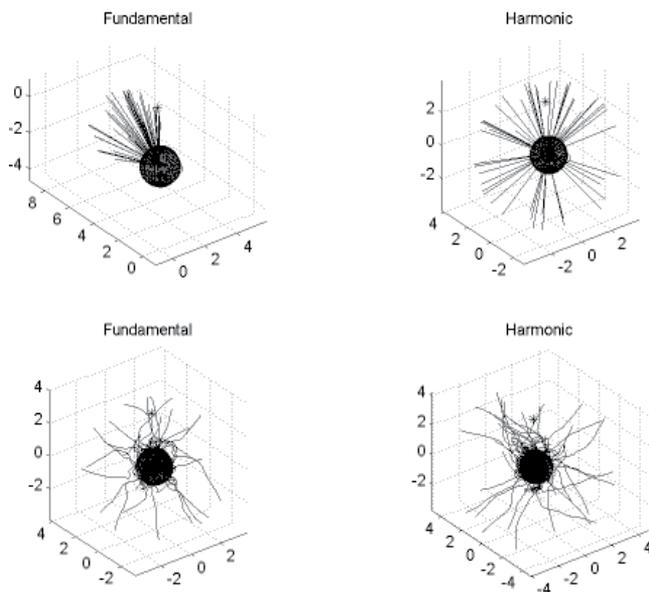


Fig. 6. The typical distributions of the refracted and scattered rays from the actual source location until they reach distances of Ulysses spacecraft. The left and right columns correspond to the fundamental and harmonic emissions, respectively. In these distribution diagrams, the locations of Ulysses and Wind are also shown as \* and o, respectively.

y direction represents the west longitude. By summing the optical depths computed at each step, the total optical depth  $\tau$  along each ray is calculated.

### 6.1 Ray trajectories

In Fig. 7, we present the typical trajectories of the traced rays. The top and bottom panels correspond to 34.5 and 73.8 MHz, respectively. The red trajectories correspond to the case, where only the refraction is considered. The refractive bending is very clear from these ray trajectories. The blue trajectories, on the other hand, correspond to the case, where refraction as well as scattering are considered. The random deflections of the rays are due to scattering by random density fluctuations. It is interesting to note that the scattered rays in the top panel (34.5 MHz) turn back before reaching the critical layer, i.e., much earlier than the refracted rays. This indicates that the scattering raises the East-West diameter of the Sun at 34.5 MHz. On the other hand, the turning points of the scattered rays in the bottom panel (73.8 MHz) almost coincide with the critical layer, similar to the refracted rays. This indicates that the scattering may not affect the East-West diameters of the radio sun at 73.8 MHz.

### 6.2 Brightness temperature distribution

To calculate the brightness temperatures for different longitudes, corresponding to different values of  $y$ , we trace fifty rays at intervals of 0.25 solar radii for each  $y$ . For each ray, we determine the brightness temperature using the computed total optical depth of the ray. Then, we average the brightness temperatures of all rays traced in that direction. By the principle of reciprocity, this value represents the brightness temperature of thermal emission from that point on the disk. The error bars are estimated by using the variance of the measures contributing to the mean. In Fig. 8, we present the distributions of the brightness temperature

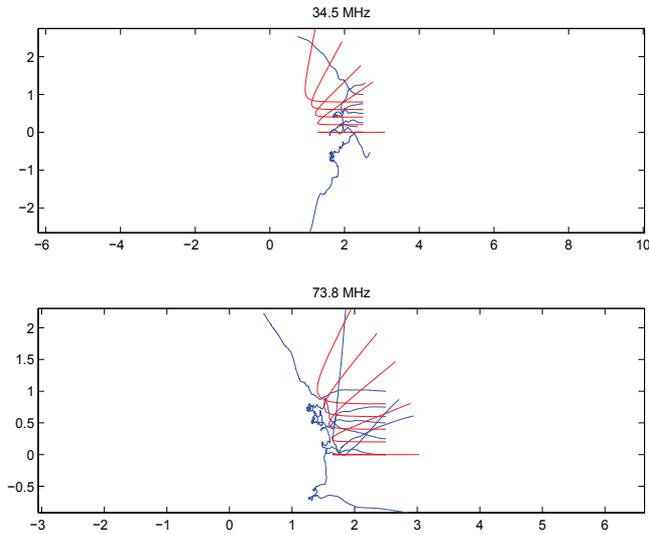


Fig. 7. Typical ray trajectories traced in the refracting (red) and refracting and scattering (blue) at 34.5 and 73.8 MHz frequencies in a non-spherical symmetric corona. The rays are initially directed toward points on the solar disk in the intervals of  $0.25R_{\odot}$ . The refractive bending is clearly seen from these trajectories. All the rays were launched along the Earth-Sun line. The random deflections of the scattered central rays are clearly visible from this figure.

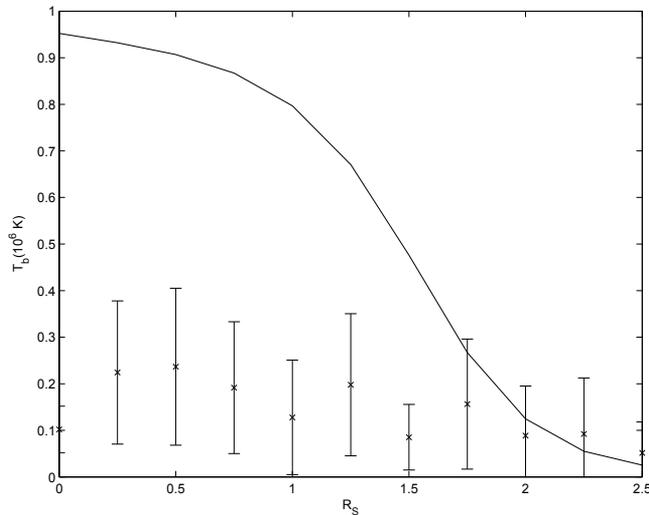


Fig. 8. Brightness temperature ( $T_B$ ) distributions for 50.0 MHz radiation. The error bars correspond to the r.m.s deviation from the mean of the computed  $T_B$  for individual rays.

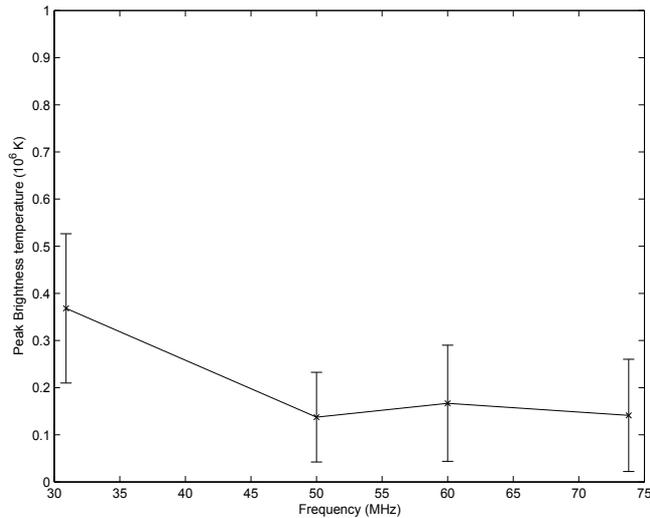


Fig. 9. Peak brightness temperature ( $T_B$ ) of the quiet Sun as a function of frequency  $f$

at 50 MHz, where the solid curve corresponds to the case, where only the refraction is considered, and the points with error bars correspond to the case, where both the refraction and scattering are included in the calculations. It is interesting to note that this brightness temperature distribution (Fig. 8) resembles very much to those computed by previous authors (Aubier et al , 1971; Riddle , 1974a; Thejappa & Kundu , 1992). We have calculated the half-power angular diameter (angular width at half-maximum) in units of arc minutes using the brightness distribution of Fig. 8, using a cubic polynomial interpolation technique. The E-W diameter of the radio sun is  $\sim 48'$  at 50 MHz when only the refraction is included. It is increased to  $\sim 56'$  when the scattering is added in the calculations. This value compares reasonably well with observed value of  $50'$  within the limits of error bars. As far as the brightness temperature is concerned, the peak value when scattering was absent was  $9 \times 10^5$  K, which is reduced to  $\sim 2.3 \times 10^5$  K when the scattering was included, i.e., the brightness temperature is reduced almost by 50% due to scattering. We have statistically computed the most probable central brightness temperatures (zero longitude, or  $y = 0$ ) for four different frequencies (30.9, 50, 60, 73.8 MHz), which are presented as the computed spectrum in fig. 9. The shape of the spectrum of thermal emission is preserved in the considered frequency range, even in the presence of density fluctuations by remaining almost steady at these frequencies in the limits of error bars. The central brightness temperatures are  $3.5 \times 10^5$  K,  $1.4 \times 10^5$  K,  $1.7 \times 10^5$  K, and  $1.4 \times 10^5$  K at 30.9, 50, 60 and 73.8 MHz, respectively.

## 7. Discussion

We have described Monte Carlo methods used in solar radio astronomy and demonstrated how to use them to explain some of the unusual characteristics of solar radio bursts as well as quiet sun radio emissions. We have examined the effects of propagation on the directivities, time profiles, sizes and positions of the 120 kHz radio burst emissions located at  $\sim 115$  kHz and 60 kHz plasma levels. The results are applicable for any fundamental (F) and harmonic (H) emissions. We have also examined to what extent the scattering is responsible for the

very low brightness temperatures and very large apparent source sizes of the quiet sun radio emissions.

### 7.1 Type III radio bursts

First, we have examined whether the widespread visibility of the fundamental and harmonic components is due to scattering. For such purpose, we have assumed that the type III bursts are emitted by isotropic point sources by ignoring the effects of finite source sizes and different shapes of the emission patterns. Even for such simplistic input parameters, the simulations have shown that the scattering increases the visibilities of the fundamental and harmonic emissions from  $18^\circ$  to  $100^\circ$ , and from  $80^\circ$  to  $150^\circ$ , respectively. The assumption of the isotropic emission patterns does affect the results because, first of all the refractive focusing does not depend on the shape of the emission patterns, and second of all the emission patterns are completely destroyed by the scattering. On the other hand, the introduction of a finite source size may increase the visibility as well as the sizes of the apparent sources.

The second question, we have examined is the effect of scattering on the time profiles of radio bursts. We have constructed the time profiles of the isotropic point sources using the arrival times of the scattered rays at 1 AU, and found them to be very similar to the observed type III burst profiles, i.e., a large fraction of the durations and exponential decays is due to propagation effects, especially due to scattering. Thus, these simulations conclusively show that we can use the durations and decay constants obtained from the observed time profiles to derive the characteristics of the electron beam and the electron temperature of the ambient plasma, only if we correct them for the propagation effects.

The third question concerns the connection between the mode of the observed emission and the propagation effects. We have shown that when the level of density fluctuations is low, the fundamental emission is dominant in a narrow range of angles around the radial direction, i.e., along the direction of the beam. Thus the identification of the mode of the observed emission as the fundamental, or a mixture of a strong fundamental and a weak harmonic by an independent technique can serve as a good indicator for the oncoming flare electrons or CME driven shock accelerated electrons at the spacecraft. On the other hand the scattering allows both the fundamental and harmonic modes to be equally visible at low longitudes, and only harmonic to be visible at high longitudes. We have also examined the question of time delay between the fundamental and harmonic emission peaks in the observed time profiles, and have shown that it depends critically on the speed of the electron beam and the position of the observer. The computations have also indicated that the usually observed single emission peak consisting of a mixture of the fundamental and harmonic emissions in a given time profile can be easily accounted for by the time delays due to propagation effects, location of the observer, and the beam speed.

The fourth question is concerned with the effects of refraction and scattering on the sizes and positions of the fundamental and harmonic emissions. We have shown that the propagation effects increase the sizes and heights of the radio sources considerably. The refraction lowers the heights of the centroids of the fundamental as well as harmonic sources, contradicting the observations that the heliocentric distances of the 100 kHz type III burst source centroids at a given frequency  $f$  range from the distance where  $f_{pe} = f/2$  to that where  $f_{pe} = f/5$ . This suggests that the refraction alone can not account for either the observed source sizes or source heights. When the scattering is included, the angular sizes are increased to  $\sim 25^\circ$  for the fundamental and  $\sim 37^\circ$  for the harmonic, and the heliocentric distances of their centroids increased to  $f_{pe} = f/3$  and  $f_{pe} \sim f/2$  levels, respectively. These values agree very well with

observations. One should note that higher value of  $\epsilon$ , for example, 0.1, may provide a still better agreement with observations, yielding larger source sizes and higher altitudes.

The distributions of the traced rays from the actual location of the source to distances of Ulysses show that the scattered fundamental, refracted harmonic as well as the scattered harmonic can easily account for the widespread visibility of radio bursts to Ulysses and Wind spacecraft. This also shows that the location of the source will have considerable influence on its visibility.

## 7.2 Quiet sun radio emission

The radio emission at meter and decameter wavelengths from the solar corona during the quiet periods of solar activity is one of the best examples of the thermal continuum emissions in nature. By measuring the brightness temperatures of this emission at different wavelengths, in principle, we can obtain the electron temperature in the corona at different heights. However, the observations have shown that the quiet sun radio emission exhibits very peculiar behavior. The East-West diameters are larger than the expected values from the thermal emission. Because of this peculiar behavior, the radio methods do not yield the correct electron temperatures. Earlier investigations (Aubier et al , 1971; Thejappa & Kundu , 1992; 1994) have shown that the scattering probably is responsible for such unusual behavior. However, in those studies, spherical symmetric models for the electron density and Gaussian power spectrum for the the density fluctuations were used. Those models do not represent the real solar conditions.

In this study, we have used the non-spherical symmetric density model for the quiet corona, which shows a bulge in equatorial region and a sort of compression in polar regions. This density distribution represents the actual observed shape of the quiet corona. Extensive observations have also shown that the spectrum of density fluctuations are correctly described by a power-law. The flat part of this spectrum with spectral index,  $\alpha = 3$  contains most of the power. In this spectral range, the geometric optics approximation is also valid. We have used a range of density scales, ranging from 50 to 75 km with axial ratio of 10. These are very close to the scale of  $5 \times 10^{-5} R_{\odot}$  used by previous authors. We have derived an expression for the angular deflection suffered by a ray due to power-law spectrum with  $\alpha = 3$  and with spatial scales of  $\frac{L_i}{L_o} \sim 0.7$  and have shown that it is almost identical to that of Gaussian spectrum. The observations also indicate that the relative level of density fluctuations,  $\epsilon = \frac{\Delta N_e}{N_e}$  is 0.1. We have used this value in the simulations. We have shown that these values cause a reduction in the brightness temperatures by almost by an order of magnitude by increasing the sizes considerably. These two results are consistent with observations. We have used for example, a value of  $T_e = 10^6$  for all the calculations. However, in order to extract an accurate information about the electron temperature  $T_e$ , from the observed brightness temperatures, one should use the models for both electron density as well as electron density fluctuations based on observations (preferably real-time), and statistically calculate the brightness temperatures for various electron temperatures at different wavelengths. The value of  $T_e$  which yields a correct value of  $T_B$  will represent the correct electron temperature at a given height. Sometimes, abnormally low brightness temperatures are observed as reported by Thejappa & Kundu (1992). These cases may be due to equatorial coronal holes with very low electron temperatures. For example, using the data from two SOHO spectrometers CDS and SUMER, David et al (1998) have shown that in a polar coronal hole, the electron temperatures are around 0.8 Mk close to the limb, rising to a maximum of less than 1 MK at  $1.15 R_{\odot}$ , then

falling around 0.4 MK at  $1.3R_{\odot}$  and 0.3 MK at  $1.6R_{\odot}$  (Wilhelm et al , 1998). In these cases, the scattering definitely can lead to very low brightness temperatures.

## 8. Conclusions

The main results from these Monte Carlo simulations are: (1) the widespread visibility of radio bursts is due to scattering of radio waves by density fluctuations, (2) the scattered fundamental and harmonic emissions produce time profiles which look very much like the idealized type III radio bursts indicating that the duration of the beam and collisional decay constants can be derived from the observed time profiles only after correcting for the propagation effects, (3) the identification of the emission modes in the type III burst time profile, namely the fundamental in the rise part followed by the harmonic in the decay part of the time profile can be accounted for by the scattering, (4) the sizes and heights of the apparent sources derived using the distributions of scattered rays from the isotropic point sources agree very well with observed values, (5) the scattering at meter-decameter wavelengths leads to a considerable reduction in the central brightness temperatures, (6) although scattering causes the reduction in the central brightness temperatures, the resultant spectrum, i.e., the peak brightness temperature as a function of the frequency remains very similar to the thermal spectrum of electromagnetic radiation, and (7) by knowing the density distribution, and the parameters of density fluctuations during the radio observations, we can determine the electron temperatures of the solar corona using the Monte Carlo simulations.

The Monte Carlo simulation methods developed in this study are very general. These techniques can be used to study the propagation of waves in any refracting and scattering medium. The diagnostics developed in this study to calculate the directivities, time profiles, sizes and positions of the radio sources can be used in a variety of contexts in solar radio astronomy. Since the scattering is probably responsible for the reduction in the intensities of thermal emission from the quiet Sun, and for the increase in the source size of the quiet Sun, a proper simulation of this process can yield an accurate determination of the electron temperature,  $T_e$ .

## 9. Acknowledgements

The research of Thejappa Golla is supported by the NASA grants NNX08AO02G and NNX09AB19G.

## 10. References

- Armstrong, J. W., Coles, W.A., Kojima, M., & Rickett, B. J. (1990), *Astrophys. J.*, 358, 685.  
Aubier, M., Leblanc, Y., & Boisshot, A. (1971), *Astron. Astrophys.*, 12, 435.  
Bale, S. D., Reiner, M. J., Bougeret, J.-L., Kaiser, M. L., Krucker, S., Larson, D. E., & Lin, R. P. (1999), *Geophys. Res. Lett.*, 26, 1573.  
Bastian, T. S., (1994), *Astrophys. J.*, 426, 774.  
Bavassano, B., & Bruno, R. (1995), *J. Geophys. Res.*, 100, 9475.  
Bougeret, J. L., Fainberg, J., & Stone, R. G. (1984a), *Astro. Astrophys.*, 141, 17.  
Bougeret, J.-L., King, J. H., & Schwenn, R. (1984b), *Sol. Phys.*, 90, 401.  
Bougeret, J.-L. et al., (1995), *Spa. Sci. Rev.*, 71, 231.  
Bracewell, R. N., & Preston, G. W. (1956), *Astrophys. J.*, 123, 14.  
Cairns, I. H. (1998), *Astrophys. J.*, 506, 456.  
Caroubalos, C., Aubier, M., Leblanc, Y & Steinberg, J. L. (1972), *Astron. Astrophys.*, 16, 374.

- Caroubalos, C., & Steinberg, J. L. (1974), *Astron. Astrophys.*, 32, 245.
- Caroubalos, C., Poquerusse, M., & Steinberg, J. L. (1974), *Astron. Astrophys.*, 32, 255.
- Chandrasekhar, S. (1952), *Mon. Noti. Roy. Astron. Soc.*, 112, 475.
- Coles, W. A., & Filice, J. P. (1985), *J. Geophys. Res.*, 90, 5082.
- Coles, W. A., Frehlich, R. G., Eickett, B. J., & Codona, J. L. (1987), *Astrophys. J.*, 315, 666.
- Coles, W. A., & Harmon, J. K. (1989), *Astrophys. J.*, 337, 1023.
- Coles, W. A., Liu, W., Harmon, J. K. & Martin, C. L., (1991), *J. Geophys. Res.*, 96, 1745.
- Coles, W. A., Rao, A. P., & Ananthakrishnana, S., (2002) *Solar Wind 10*, Pisa.
- David, C., Gabriel, A. H., Bely-Dubau, F., Fludra, A., Lemaire, P., & Wilhelm, K., (1998), *Astron. Astrophys.*, 336, L90.
- Dulk, G. A., Steinberg, J. L., & Hoang, S. (1984), *Astro. Astrophys.*, 141, 30.
- Dulk, G. A., Steinberg, J.-L., Lecacheux, A., Hoang, S., & MacDowall, R. J. (1985), *Astron. Astrophys.*, 150, L28.
- Dulk, G. A., Leblanc, Y., Bougeret, J. L., & Hoang, S. (1996), *Geophys. Res. Lett.*, 23, 1203.
- Efimov, A. I., Chashei, I. V., Bird, M. K., Samoznaev, L. N., & Plettemeier, D. (2005), *Astro. Rep.*, 49, 485.
- Fainberg, J., & Stone, R. G. (1970), *Sol. Phys.*, 15, 222.
- Fainberg, J., Evans, L. G., & Stone, R. G. (1972), *Science*, 178, 743.
- Fainberg, J., & Stone, R. G. (1974), *Spa. Sci. Rev.*, 16, 145.
- Fitzenreiter, R. J., Fainberg, J., Bundy, R. B. (1976), *Sol. Phys.*, 46, 465.
- Fokker, A. D. (1965), *Bul. Astro. Inst. Neth.*, 18, 111.
- Ginzburg, V. L., & Zheleznyakov, V. V. (1958), *Sov. Astro.*, 2, 653.
- Golap, K., & Sastry, Ch. V., (1994), *Sol. Phys.*, 295, 150.
- Grall, R. R., Coles, W. A., Spangler, S. R., Sakurai, T., & Harmon, J. K., (1997), *J. Geophys. Res.*, 102, 263.
- Guhathakurta, M., Holzer, T. E., & MacQueen, R. M., (1996), *Astrophys. J.*, 458, 817.
- Gurnett, D. A., & Anderson, R. R. (1976), *Science*, 194, 1159.
- Gurnett, D. A., Baumbach, M. M., & Rosenbauer, H. (1978), *J. Geophys. Res.*, 83, 616.
- Haddock, F. G., & Graedel, T. F. (1970), *Astrophys. J.*, 160, 293.
- Hartz, T. R. (1969), *Plan. Spac. Sci.*, 11, 115.
- Haselgrove, J. (1963), *J. Atmos. Terr. Phys.*, 25, 397.
- Hoang, S., & Steinberg, J. L., (1977) *Astro. Astrophys.*, 58, 287.
- Hoang, S., Maksimovic, M., Bougeret, J.-L., Reiner, M. J., & Kaiser, M. L. (1998), *Geophys. Res. Lett.*, 25, 2497.
- Hollweg, J. (1968), *Astron. J.*, 73, 972.
- Hughes, M. P., & Harkness, R. L. (1963), *Astrophys. J.*, 138, 239.
- Jaeger, J. C., & Westfold, K. C. (1950), *Austr. J. Res.*, (A), 2, 322.
- Kaiser, M. L. (1975), *Sol. Phys.*, 45, 181.
- Kellogg, P. J. (1980), *Astrophys. J.*, 236, 696.
- Kundu, M. R., (1965), *Solar Radio Astronomy*, Interscience Publishers.
- Lacombe, C., Harvey, C. C., Hoang, S., Mangeney, A., Steinberg, J.-L., & Burgess, D. (1988), *Ann. Geophys.*, 6, 113.
- Lantos, P., (1980) in *Radiophysics of the Sun* (M. R. Kundu and T. Gergeley eds.), Proceedings of IAU Symposium No. 86.
- Lantos, P., (1998), *Solar Physics with Radio observations*, proc. Nobeyama Symposium, NRO Report 479.
- Leblanc, Y. (1973), *Astrophys. J. Lett.*, 14, 41.

- Lecacheux, A., Steinberg, J.-L., Hoang, S., & Dulk, G. A. (1989), *Astron. Astrophys.*, 217, 237.
- Lee, L. C., & Jokipi, J. R. (1975), *Astrophys. J.*, 196, 695.
- Lin, R. P., Levedahl, W. K., Lotko, W., Gurnett, D. A., & Scarf, F. L. (1986), *Astrophys. J.*, 308, 954.
- MacDowall, R. J. (1983), *M. S. Thesis*, University of Maryland.
- Manoharan, P. K., Ananthakrishnan, S., & Rao, A. P. (1988), *Proc. Sixth International Solar Wind Conf. Vol. 1* (Boulder, NCAR), 55.
- Narayan, R., Anantharamaiah, K. R., & Cornwell, T. J. (1989), *Mon. Not. Roy. Astron. Soc.*, 241, 403.
- Newkirk, G. A., (1961), *Astrophys. J.*, 133, 983.
- Poquerusse, M., Steinberg, J. L., Caroubalos, C., Dulk, G. A., & MacQueen, R. M. (1988), *Astron. Astrophys.*, 192, 323.
- Ramesh, R., Nataraj, H. S., Kathiravan, C., & Sastry, Ch. V. (2006), *Astrophys. J.*, 648, 707.
- Reiner, M. j., & Stone, R. G. (1988), *Astron. Astrophys.*, 206, 316.
- Reiner, M. j., & Stone, R. G. (1989), *Astron. Astrophys.*, 217, 251.
- Reiner, M. J., Fainberg, J., & Stone, R. G. (1992), *Astrophys. J.*, 394, 340.
- Reiner, M. J., Fainberg, J., Kaiser, M. L., & Stone, R. G. (1998), *J. Geophys. Res.*, 103, 1923.
- Rickett, B. J. (1977), *Ann. Rev. Astron. Astrophys.*, 15, 479.
- Riddle, A. C. (1972), *Proc. Astron. Soc. Austr.*, 1972,2, 98.
- Riddle, A. C. (1974a), *Sol. Phys.*, 36, 375.
- Riddle, A. C. (1974b), *Sol. Phys.*, 35, 153.
- Rytov, S. M., Kravtsov, Yu. A., & Tatarskii, V. I. (1989), *Principles of Statistical Radiophysics. vol. 4. Wave Propagation Through Random Media*. Springer-Verlag.
- Sastry, Ch. V. (1994), *Sol. Phys.*, 150, 285.
- Sawyer, C., & Warwick, J. W. (1987), *Astron. Astrophys.*, 206, 316.
- Sharma, A., Vizia Kumar, D., & Ghatak, A. K. (1982), *Appl. Opt.*, 21.
- Sheridan, K. V., & McLean, D. J. (1985), (in *Solar Radiophysics* (D. J. McLean and N. R. Labrum, eds), Cambridge University press, Cambridge.
- Smerd, S. F. (1950), *Australian J. Sc., Res.*, A3, 34.
- Smith, D. F. (1970), *Sol. Phys.*, 15, 202.
- Spangler, S. R., & Sakurai, T. (1995), *Astrophys. J.*, 445, 999.
- Spangler, S. R., Kavars, D. W., Kortenkamp, P. S., Bondi, M., Mantovani, F., & Alef, W. (2002), *Astron. Astrophys.*, 384, 654.
- Spangler, S. R. (2002), *Astrophys. J.*, 576, 997.
- Steinberg, J.-L., Aubier-Giraud, M., Leblanc, Y., & Boisshot, A. (1971), *Astro. Astrophys.*, 10, 362.
- Steinberg, J. L. (1972), *Astro. Astrophys.*, 18, 382.
- Steinberg, J.-L., Dulk, G. A., Hoang, S., Lecacheux, A., & Aubier, M. G. (1984), *Astro. Astrophys.*, 140, 39.
- Steinberg, J.-L., Hoang, S., & Dulk, G. A. (1985), *Astro. Astrophys.*, 150, 205.
- Stone, R. G., et al. 1992, *Astron. Astrophys. Supp.*, 92, 291.
- Subramanian, K. R. (2004), *Astro. Astrophys.*, 426, 329.
- Tarnstrom, G. I., & Philip, K. W. (1972), *Astro. Astrophys.*, 17, 267.
- Thejappa, G., & Kundu, M. R. (1992), *Sol. Phys.*, 140, 19.
- Thejappa, G., Lengyel-Frey, D., Stone, R. G., & Goldstein, M. L. (1993), *Astrophys. J.*, 416, 831.
- Thejappa, G., & Kundu, M. R. (1994), *Sol. Phys.*, 149, 31.
- Thejappa, G., & MacDowall, R. J. (1998), *Astrophys. J.*, 498, 465.

- Thejappa, G., MacDowall, R. J., & Kaiser, M. L. (2007), *Astrophys. J.*, 671, 894.
- Thejappa, G., & MacDowall, R. J. (2008a), *Astrophys. J.*, 676, 1338.
- Thejappa, G., & MacDowall, R. J. (2008b), (in) *Turbulence, Dynamos, Accretion Disks, Pulsars and Collective Plasma processes*, Astrophysics and Space Science proceedings, part VI, 311-328, DOI:10.1007/978-1-4020-8826-1-21.
- Thejappa, G., & MacDowall, R. J. (2010), *Astrophys. J.*, 720, 1395.
- Tu, C. Y., & Marsch, E. (1994), *J. Geophys. Res.*, 9921,481.
- Wilhelm, K., Marsch, E., Dwivedi, B. N., Hassler, D. M., Lemaire, P., Gabriel, A. H., & Huber, M. C. E. (1998), *Astrophys. J.*, 500, 1023.
- Wohlmuth, R., Plettemeier, D., Edenhofer, P., Bird, M. K., Efimov, A. I., Andreev, V. E., Samoznaev, L. N., & Chashei, I. V. (2001), *Spa. Sci. Rev.*, 97, 9.
- Woo, R., Armstrong, J. W., Bird, M. K., & Patzold, M. (1995), *Geophys. Res. Lett.*, 22, 329.
- Zheleznyakov, V. V., & Zaitsev, V. V. (1970), *Sov. Astron.*, 14, 250.

# Using Monte Carlo Simulation for Prediction of Tool Life

Sayyad Zahid Qamar<sup>1</sup>, Anwar Khalil Sheikh<sup>2</sup>,  
Tasneem Pervez<sup>1</sup> and Abul Fazal M. Arif<sup>2</sup>

<sup>1</sup>Mechanical and Industrial Engineering Department,  
Sultan Qaboos University, Muscat,

<sup>2</sup>Mechanical Engineering Department,  
King Fahd University of Petroleum and Minerals, Dhahran,  
<sup>1</sup>Oman

<sup>2</sup>Saudi Arabia

## 1. Introduction

Hot metal forming (rolling, forging, extrusion, wire drawing) constitutes a very large proportion of manufacturing activity. Of all the equipment and tooling involved in a hot forming process, the most critical component is usually considered to be the die due to its superior precision and reliability requirement and the associated high cost. Dies and ancillary tooling are exposed to high pressures, elevated temperatures, and both mechanical and thermal fatigue. Cost and engineering difficulty are obviously high because of factors such as special material and processing, very fine tolerances, and high demands on repeated thermo-mechanical performance. How often a die has to be scrapped and replaced with a new one directly contributes to the commercial viability of producing a certain profile, not only because of the large cost of die replacement, but also because of losses due to interrupted production and reduced product quality in the case of a defective die.

Due to this critical importance and high cost of metal forming tools, one of the major goals is a longer tool life. Continued research in tool and process design is therefore targeted at minimization of tool failure. However, tool failure is a complex phenomenon, governed by an interaction of various mechanisms, and not easy to control or restrain. A more manageable approach is to make die failure *predictable*. Estimation and prediction of tool life thus become crucially important.

Extrusion is one of the most popular metal forming processes. Because of its wide-ranging and abundant application in the automobile, aircraft, and construction industries, aluminum has been called the *metal of the millennium*. As extrusion is the primary manufacturing process for aluminum alloys, the popularity and importance of extrusion has increased even more. The three leading failure mechanisms for extrusion dies are fracture, wear, and plastic deformation (Bauser et al., 2006). Fatigue fracture and gradual wear are the more dominant of these failure modes.

### 1.1 Fatigue fracture

As established in an earlier work by the authors (Arif et al., 2003), fracture is the principal failure mode for extrusion dies and tooling (solid, hollow, and semi-hollow die profiles all taken together). An extrusion die experiences both mechanical and thermal stresses during its service life. As temperature changes are rather gradual, thermal stresses due to temperature differences are generally not very critical. Mechanical stresses are cyclic in nature, going from zero to maximum and back during extrusion of each billet. The maximum stress in aluminum extrusion can be around 27,000 psi or 186 MPa (Laue and Stenger, 1981). At the beginning of each extrusion cycle (after loading a fresh billet into the container), the pressure applied by the ram of the extrusion press swiftly reaches a maximum value, the operation being known as *upsetting*. As extrusion proceeds, the pressure gradually decreases to an approximately constant value; the *steady-state* portion of the process. Towards the end of each cycle, pressure may increase again, during butt removal or *discard rejection* (removal of the end portion of the billet). These cyclic stresses, in the presence of some pre-existing flaws (such as micro cracks produced during heat treatment), can lead to crack growth and ultimate fatigue failure. The failure mechanism is influenced by

- a. material properties, geometric tolerances, and surface finish of the billet;
- b. material properties, heat treatment and surface hardening, and geometrical details of the die and tooling;
- c. stress distribution and variation with time and temperature during the extrusion process; and
- d. stiffness and kinematics of the press and affiliated tooling.

Fatigue failures are mostly located at positions of high stress concentration such as sharp corners, section changes, stamp marks, etc.

Fig. 1 shows schematic and actual occurrence of the significant failure mechanisms for an extrusion die. The incidence of *fatigue fracture* is especially high for high extrusion ratios and small fillet radii, resulting in high stress concentration. Crack initiation may be further promoted by machining marks. Failure by *forced rupture* (due to overload) rarely occurs in practical applications, and can often be traced to human errors. Some aspects of fatigue and fracture failure in extrusion dies and tooling have been investigated by different researchers (Hambli and Badie-Levet, 2000; Gouveia et al., 2000; Sudhakar, 2002; Yoh et al., 2002; Cosenza et al., 2004; Tseronis et al., 2008; and Nanninga et al., 2009).

### 1.2 Surface wear

The second most significant failure mode in extrusion dies is gradual wear of the die bearing surface (Arif et al., 2003). A combination of factors such as intricate profile geometries, high pressures and temperatures, very hard die material, and extremely hard and abrasive surface layer of  $Al_2O_3$  formed on the billet surface during preheating lead to wear at the die land. Die wear is a tribological effect and can be defined as the progressive loss or removal of material from the operating surface of tooling components (die bearing surface in our case). By changing the topography of the die land, wear can cause severe surface damage leading to product defects and finally die failure.

Both adhesive and abrasive wear mechanisms work together with sudden temperature fluctuations and prolonged exposures to elevated temperatures. *Abrasive wear* is more gradual, but is quite accelerated at high temperatures. Resulting die wash (wear) is at times aggravated by *adhesive wear*. It was reported by Thedja et al. (1992) that die wear begins on

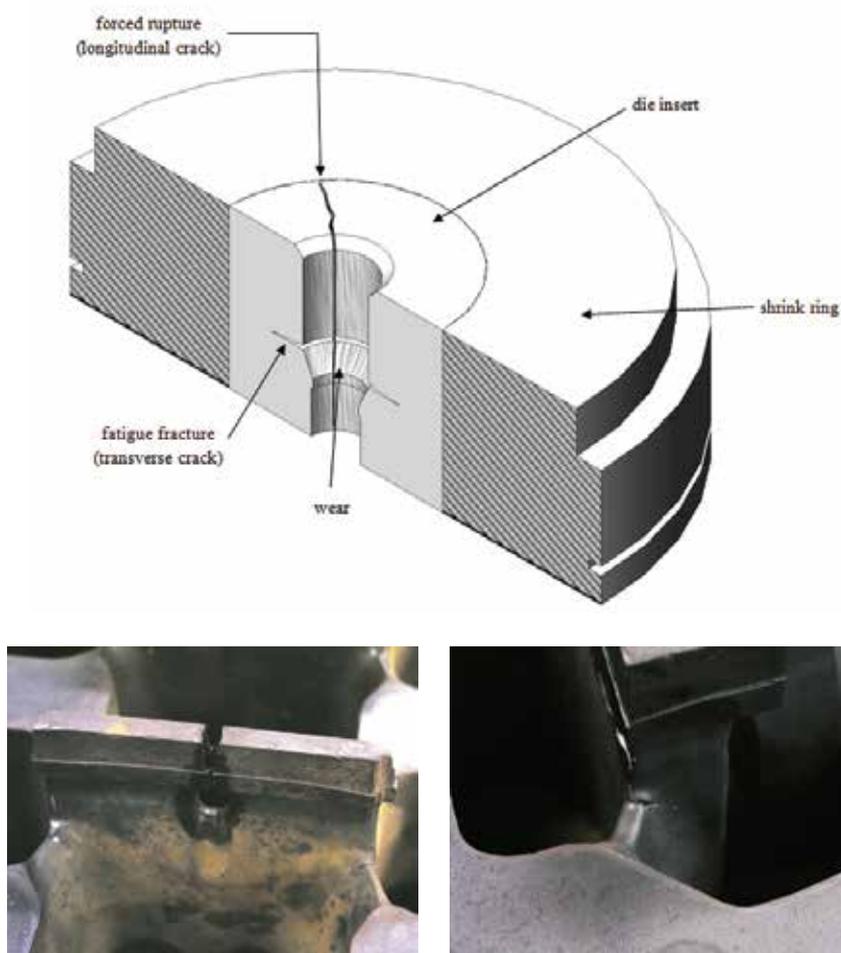


Fig. 1. Schematic and actual illustrations of fracture failure of dies in extrusion

the output side of the die bearing and progresses in a direction opposite to that of extrusion. According to other studies, the deepest wear traces can also be found at the leading edge (Saha, 1998) or in the middle (Björk et al., 2001). The disagreement in these findings is most likely caused by differences in extrusion conditions, bearing parameters, profile geometries, etc. Saha (1998) observed significant bearing washout (die wear) on a hollow die studied, and found that the wear increased at higher extrusion speeds. More wear spots were found on the mandrel bearing surface than on the cap bearing. Wear patterns observed by Thedja et al. (1992) in a hollow die are shown in Fig. 2. More detailed description of some aspects of die wear in extrusion and other metal forming dies can be found in Lee and Im (1999), Björk et al. (2001), Müller (2002), Terčelj et al. (2007), and So et al. (2008).

### 1.3 Current work

As mentioned earlier, the two most dominant failure mechanisms for extrusion dies are fracture and wear. Work presented here focuses on the development of life prediction models for extrusion dies based on fracture, wear, and combined fracture-wear failure

mechanisms. In the first part of this chapter, a fracture mechanics based fatigue life prediction model is presented. A similar treatment is then presented for wear-related failures. Fracture and wear usually coexist as failure modes, and final die breakdown occurs due to the mechanism that becomes dominant. Therefore, a competing fracture-wear model has been later developed to represent the complete die failure situation.

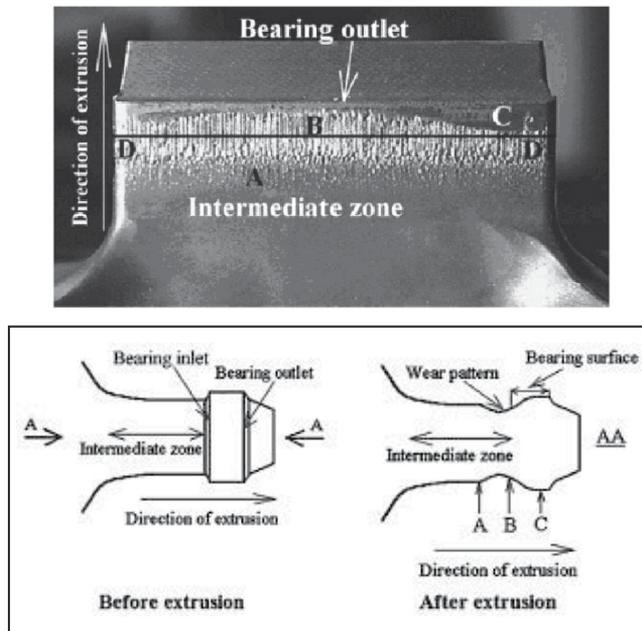


Fig. 2. Wear pattern in the mandrel of a hollow die (above); sketch showing mandrel before and after extrusion (below); longitudinal sectional view of the mandrel shows excessive wear (Thedja et al., 1992)

A probabilistic rather than a deterministic approach has been adopted (far closer to real behavior), with an attempt to correlate the stochastic nature of various fatigue and wear related die parameters to die life. Geometric features of the billet and tooling, their material properties, and relevant process parameters (extrusion pressure, ram speed, etc) are treated as random quantities. Nature of distributions and related parameters (mean, standard deviation, etc) for the different variables are determined from experimental and manufacturing data. The well-established Monte Carlo method is used to simulate instances of fracture, wear, and combined fracture-wear failures of the die for a given set of manufacturing conditions and mechanical properties. Case studies of actual hot aluminum extrusion from the industry are used for model development and validation. Two die profiles (a tube die and a box die) have been included, which are simple in geometry, but represent shape complexity to some degree, as hollow dies have a more complex extrusion setup than solid dies. The work can be extended to cover more complex profiles, thus providing life prediction for a die of a given profile *before it is put to use*. This life prediction can be very useful for developing *optimum die replacement strategies* in the industry, reducing warehousing costs significantly, and guarding against unnecessary downtime due to unavailability of a certain die profile. It can also contribute to *improvement of die design* by finding out expected failure times due to fracture and wear.

## 2. Case study

Actual die failure data have been collected from a typical medium-to-large size commercial aluminum extrusion setup for two simple hollow dies. The tube die predominantly failed by fracture, and only a few times by wear; outer diameter  $D_o = 25.4 \pm (0.2, 0.1)$  mm and thickness is  $t = 1.6 \pm (0.15, 0.1)$  mm. This die has been used to simulate fracture, wear, and combined fracture-wear failure mechanisms. The box die actually failed almost entirely due to die land wear; it is therefore used only for parameter estimation of wear failures; length of 40 mm, width of 20 mm, and thickness of  $1.3 \pm (0.00, 0.15)$  mm. Die material for both the dies was heat treated and surface hardened H13 steel, billet material being the soft grade aluminum alloy Al-6063 most commonly used in the construction sector. Average extrusion temperature was around 460°C, and ram speed was 5 mm/s. The average die life (mean time to failure MTTF) for the tube die was 722 extrusion cycles, based on actual fracture failure data summarized in Table 1. Extrusion of each billet is considered to be one cycle, as the extrusion pressure goes from a minimum of zero to a maximum value, and then back to zero for each billet.

Failure #	1	2	3	4	5	6	7
Cycles to Failure	920	671	712	902	574	652	623

Table 1 Various instances of number of billets/cycles to failure (due to fracture) for the tube die studied

## 3. Fatigue failure

### 3.1 Crack initiation

In extrusion dies and tooling, cracks generally initiate at positions of high stress concentration such as section changes, sharp corners, stamp marks, etc. Crack initiation may be further promoted by machining marks. Due to manufacturing operations such as spark erosion (electric discharge machining EDM), there may be preexisting cracks, such as those shown in Fig. 3 (Pöhlandt and Kuehl, 1989). Depths of such cracks are typically about 0.01 mm.

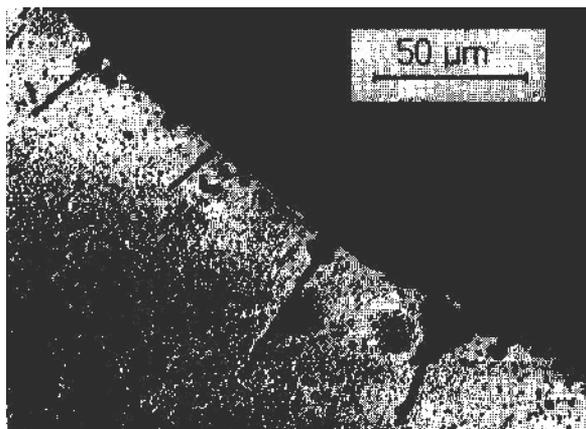


Fig. 3. Formation of fatigue cracks at the surface defects of spark eroded Cr-Mo-V steel (Pöhlandt, 1989)

To improve wear resistance, die bearing surfaces are surface hardened, usually by nitriding. This nitriding process often results in the formation of small cracks of the order of 0.05 mm for H13 tool steels (Laue and Stenger, 1981). To achieve the desired combination of hardness and toughness characteristics, dies are subjected to specific heat treatment routines (hardening/austenitizing, tempering, quenching, etc). Existing small flaws usually get enlarged during these operations. With such initial pre-operation cracks (crack depths of 0.05 to 0.1 mm), crack propagation takes over due to fatigue cycles during the actual extrusion process.

### 3.2 Crack propagation

Die fracture can be caused by one of two mechanisms. In the presence of an existing defect, a crack may take only a few cycles to reach the critical size, resulting in forced rupture (Fig. 1). This is known as *hypercritical*, instable crack growth. The other mechanism is *subcritical*, stable crack growth: from an existing defect, crack growth is caused by fatigue or creep (or a combination of the two), resulting in ultimate fracture. Both of these mechanisms are affected by different factors concerning the billet and die material, the process and the equipment. In hot extrusion, fatigue is the major contributor, effects of creep being negligible (Pöhlandt and Kuehl, 1989).

For components having subcritical cracks, fracture is almost of no concern as long as crack size is below the critical value under the given conditions. Under cyclic loading, crack propagation takes place until the stress intensity factor equals the fracture toughness of the die material. Large cyclic stresses, combined with regions of high stress concentration in cavities, lead to crack growth and ultimate failure. In general, cracks can grow only under tensile stresses, compressive stresses promoting crack closure and reduction of crack growth.

Three stages of crack growth are identified in fracture mechanics (Anderson, 2005). The first stage is the *crack initiation* or Stage-I crack growth. The second part is the Stage-II *crack growth* where a crack grows gradually and subcritically. Finally, when the fatigue crack grows to a length approaching the *critical length* for instability at the maximum applied stress level, the crack grows in an unstable fashion; the Stage-III crack growth region. Experimental studies show that fatigue crack growth data plot as a straight line on a log-log scale over Stage-II for a wide variety of metallic materials. Crack growth rate  $da/dN$ , as a function of the applied stress intensity range  $\Delta K$ , can be approximated by the famous Paris law proposed by Paris and Erdogan:

$$da / dN = C(\Delta K)^m. \quad (1)$$

Value of  $C$  depends on the system of units used for  $da/dN$  (mm/cycle or in/cycle) and  $\Delta K$  (ksi $\sqrt{\text{in}}$  or MPa $\sqrt{\text{m}}$ ). For instance, for austenitic stainless steel, Barsom and Rolfe (1970) report  $m = 3.25$  and  $C = 5.60(10^{-2})$  in mks and  $3.0(10^{-10})$  in ips units. Values of the Paris constants  $C$  and  $m$  for a number of metals have been reported by Sanford (2003) and others. Influence of non-zero average stress has been neglected in this study.

### 3.3 Fatigue life or cycles to failure

It is well-known that (neglecting the finite size factor  $f(a/W)$  for simplicity),

$$\Delta K = \alpha \Delta \sigma \sqrt{\pi a}, \quad (2)$$

where  $\Delta\sigma = \sigma_{\max} - \sigma_{\min}$ . Since each extrusion cycle starts from a minimum load of zero,  $\sigma_{\min} = 0$ . Thus,  $\Delta\sigma = \sigma_{\max}$ . Substituting equation (2) into (1), and rearranging, we get

$$dN = \frac{da}{C(\alpha\sigma_{\max}\sqrt{\pi a})^m}. \quad (3)$$

The fatigue life (or number of cycles to failure) can now be obtained by integrating this equation:

$$N_f = \frac{(a_0)^{1-m/2} - (a_c)^{1-m/2}}{C(m/2-1)\alpha^m\pi^{m/2}\sigma_{\max}^m}. \quad (4)$$

Values of  $C$  and  $m$  can be found for ultrahigh strength steels (hot-extrusion die steel H13 falls under this category) from standard references. Size of preexisting cracks ( $a_0$ ) in heat treated and surface hardened H13 steel, as mentioned above, is generally in the 0.05-0.1 mm range. Value of the geometry factor ( $a$ ) for an edge crack is 1.12.

To find the crack size that would trigger an unstable crack growth, we start from the definition of the mode-I stress intensity factor

$$K_I = \alpha f(a/W)\sigma\sqrt{\pi a}. \quad (5)$$

Neglecting the finite-size factor  $f(a/W)$ , and knowing that  $a = a_c$  when  $K_I = K_{IC}$ , we get

$$a_c = \frac{1}{\pi} \left( \frac{K_{IC}}{\alpha\sigma_{\max}} \right)^2. \quad (6)$$

For the simple case of a tube die, treating it as a thick-walled cylinder with internal pressure, the maximum stress would be (Budynas and Nisbett, 2008)

$$\sigma_{\max} = p \left( \frac{r_o^2 + r_i^2}{r_o^2 - r_i^2} \right), \quad (7)$$

where  $r_o$  and  $r_i$  are the outer and inner radii of the tube. Extrusion pressure  $p$ , considering friction at the billet-container interface but neglecting the relatively small billet-die friction, is given by Groover (2010)

$$p = \bar{Y}_f (\varepsilon + 2L/D_b). \quad (8)$$

Here, billet diameter is  $D_b$ , average flow stress of the billet material is  $\bar{Y}_f$ , and instantaneous billet length is  $L$ . At the beginning of the stroke (when a fresh billet is loaded into the container),  $L = L_0$ ; at the end of the stroke,  $L = L_b$ , where  $L_b$  is the length (thickness) of the butt remaining in the container (cut away by a shearing mechanism and removed before loading the next billet into the container). The true strain is given by

$$\varepsilon = \ln R. \quad (9)$$

Extrusion ration  $R$  can be expressed as

$$R = \frac{A_b}{n_1 A_s} = \frac{D_b^2}{n_1 (d_o^2 - d_i^2)}, \quad (10)$$

where  $n_1$  is the number of cavities in a multi-cavity die. Of course, at  $L = L_0$  (maximum value of  $L$  at the start of the extrusion cycle),  $p = p_{max}$  in equation (8). The inner diameter of the tube is

$$d_i = d_o - 2t. \quad (11)$$

Flow stress of the material can be evaluated from the relation (Saha, 2000)

$$\bar{Y}_f = \bar{\sigma} = \bar{\sigma}_0 \left( \frac{\dot{\epsilon}}{\dot{\epsilon}_0} \right)^{m^*}, \quad (12)$$

where  $\bar{\sigma}_0$  is the known flow stress at a known average strain rate  $\dot{\epsilon}_0$ . A typical value of the exponent  $m^*$  at 500°C for Al-Mg-Si alloy (the category to which 6063 and other 6xxx alloys of aluminum belong) is 0.125. At an average strain rate of 50 s<sup>-1</sup>, the average flow stress for Al-6063 was found to be around 40 MPa using the graph shown in Fig. 4 (Sheppard, 1999). The mean strain rate can be found from

$$\dot{\epsilon} = \frac{6V}{D_b} \ln R, \quad (13)$$

where  $V$  is the ram speed and  $R$  is the extrusion ratio.

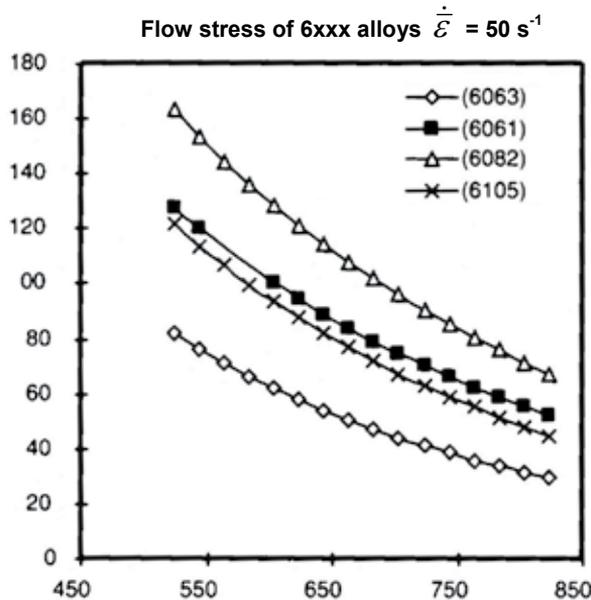


Fig. 4. Flow stress at elevated temperature for 6xxx aluminum alloys, including strain rate effect (Sheppard, 1999)

### 3.4 Probabilistic nature of die life

Fatigue life and fracture mechanics based models for life prediction are generally deterministic in nature. On the other hand, geometric parameters of an extrusion die and experimentally determined material properties are random variables. Life of an extrusion die is thus largely probabilistic in character. It has been assumed here that geometric dimensions of the die profile and the billet (outer diameter  $d$  and thickness  $t$  of the tube, length  $L$  and diameter  $D_b$  of the billet) and initial crack size  $a_0$  (preexisting flaws due to heat treatment and surface hardening) are normally distributed random variables. Mean ( $\mu$ ) and standard deviation ( $\sigma$ ) values for profile dimensions are derived from tolerances specified on manufacturer's profile drawing, and those for initial crack size are based on studies about preexisting cracks due to spark erosion in tool steels (Bauser et al., 2006). The data is assumed to be spread within  $\pm 3\sigma$  limits of the mean.

Plane strain fracture toughness  $K_{IC}$  of the die material is a strength property and is well represented by a Weibull distribution. Values of  $\mu$  and  $\sigma$  and the Weibull parameters  $m$  and  $\theta$  are based on variation in  $K_{IC}$  values of tool steels as reported in references such as Sanford (2003) and tool steel manufacturers. Paris constant  $C$  of the die material, due to the nature of the Paris law equation, has a log-normal distribution. Mean, variance, and lognormal parameter values ( $y_0$  and  $\omega$ ) are estimated from spread of crack growth data for tool steels, shown in Fig. 5 (Callister, 2006). The other quantities (Paris constant  $m$ , geometry factor  $a$ , flow stress exponent  $m^*$ , ram speed  $V$ , number of cavities  $n_1$ ) have been treated as constants. Information about the distribution type, average values, standard deviations, etc of all the variables is listed in Table 2.

Variable	Distribution	Mean Value	Std Deviation
Billet dia $D_b$	Normal	184 mm	0.5 mm
Billet length $L$	Normal	660 mm	4 mm
Die profile outer dia $d_o$	Normal	25.4 mm	0.1 mm
Die thickness $t$	Normal	1.6 mm	0.05 mm
Paris constant $C$	Lognormal	$1.6 \times 10^{-12}$ ( $y_0 = 1.493 \times 10^{-12}$ )	$0.192 \times 10^{-12}$ ( $\omega = 0.229$ )
Paris exponent $m$	Constant	2.85	-
Fracture toughness $K_{IC}$	Weibull	$83.6 \text{ MP}\sqrt{\text{m}}$ $\theta = 89.6 \text{ MPa}\sqrt{\text{m}}$	$12.54 \text{ MPa}\sqrt{\text{m}}$ $m = 6.67$
Ram speed $V$	Constant	5 mm/s	-
Initial crack size $a_0$	Normal	0.01 mm	0.001 mm
Geometry factor $a$	Constant	1.12	-
Number of cavities $n_1$	Constant	4	-

Table 2. Variables involved, their distributions and values

Usually only room temperature fracture toughness values of H13 tool steel are available for a few temper conditions (Saha, 2000). Optimum heat treatment (to obtain the best hardness and toughness combination) for H13 steel being used for hot work dies is tempering to around 550°C (Bauser et al., 2006). Average operating temperature in commercial aluminum

extrusion is about 460°C. Qamar et al. (2006) developed linear and quadratic polynomial models for the prediction of  $K_{IC}$  values for H13 steels subjected to different tempering routines and used at different operating temperatures. The quadratic model gives a higher correlation coefficient, but the linear model gives more conservative predictions:

$$\frac{K_{IC}(T)}{HRC(T)} = 2.39 \left( \frac{CVN(T)}{HRC(T)} \right) + 0.17 \quad (\text{MPa}\sqrt{\text{m}}, \text{J}) \quad (14)$$

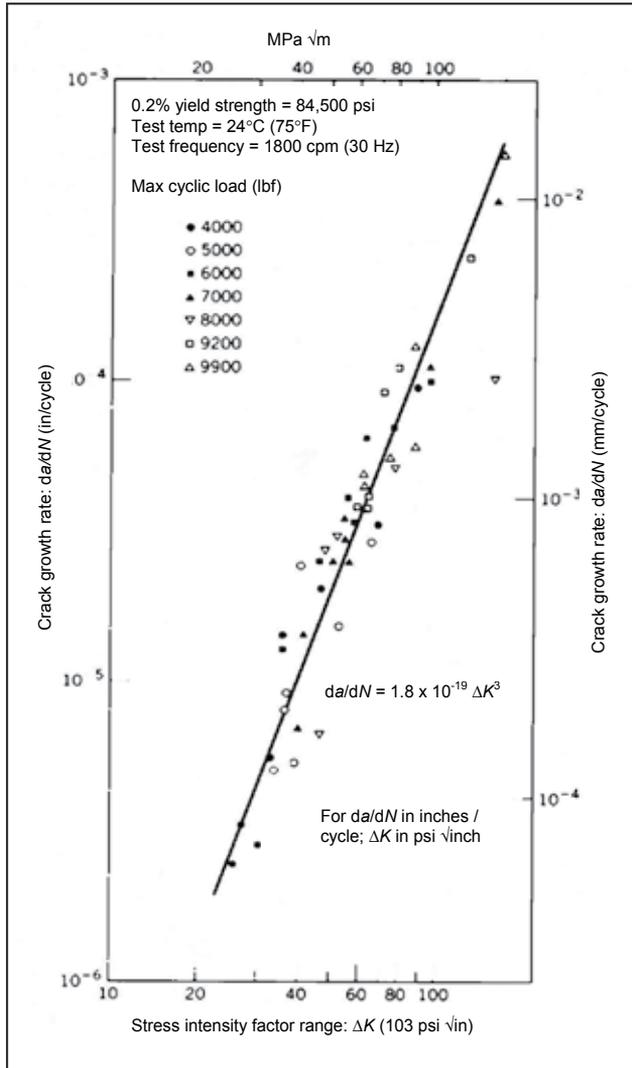


Fig. 5.  $da/dN$  vs  $\Delta K$  graph (log-log scale) for a Ni-Mo-V steel (Callister, 2006)

Based on hot hardness  $HRC(T)$  and hot impact strength  $CVN(T)$  data, the predicted fracture toughness value for the temper condition and the working temperature in our case comes out to be around 83.6  $\text{MPa}\sqrt{\text{m}}$ .

### 3.5 Monte Carlo simulation — fatigue

The beauty of the Monte Carlo method is that, if a reasonably accurate mathematical model (equation) of a physical process is available, a real experiment can be simulated as many times as required, and the randomness of the variables is guaranteed. The strategy is based upon first establishing a cause-and-effect relationship (such as equations 4 to 13) of the type

$$N = \varphi(X_1, X_2, \dots, X_N), \quad (15)$$

where  $N$  is the tool life (cycles to failure) and  $X_1, X_2, \dots, X_N$  are the various geometrical features and material properties of the die or tooling. Since  $X_1, X_2, \dots, X_N$  are random, the tool life  $N$  will also be random in nature.

A large sample (10,000 in our case) of random values is then created through generation of independent random numbers ( $Z_i$  or  $U_i$ ) using a random number generator (based on a standard normal distribution for normal and lognormal type variables, and a uniform distribution for Weibull type data). These random numbers are then transformed into the required statistical distributions through appropriate transformations. For normally distributed variables such as  $D_b, d_o, t$ , etc

$$X_i = \mu_i + \sigma_i Z_i; \quad (16)$$

for variables such as the Paris constant  $C$ , having a log-normal distribution

$$X_i = \exp(\mu_i + \sigma_i Z_i); \quad (17)$$

and for Weibull-distributed data such as  $K_{IC}$

$$X_i = \theta [\ln(1/U_i)]^{1/m}. \quad (18)$$

In these transformation equations,  $\mu$  is the mean and  $\sigma$  is the standard deviation of the normal variables based on actual data;  $\mu_i$  and  $\sigma_i$  are the mean and standard deviations of the log-normal data set; while  $m$  and  $\theta$  are the shape and scale parameters of the Weibull data (Table-2). Once all the basic variables are randomly generated in this manner, the derived variables (such as  $a_c, \sigma_{max}, p, \varepsilon$ , etc) are calculated for all the 10,000 instances. Cycles to failure (number of billets) due to fatigue fracture are of course calculated for each set of simulated variables using equation (4).

### 3.6 Die life distribution — fatigue

Reliability is the term used to express the “probability that a component or system will perform satisfactorily for a specified period of time ( $t$ ) under a given set of working conditions,” expressed by  $R(t) = P\{t > t\}$ . The associated cumulative probability is the “probability that failure takes place at a time less than or equal to  $t$ ,” expressed as  $F(t) = P\{t \leq t\}$ . It is obvious that  $F(t) = 1 - R(t)$ . For reliability characterization of life of hot working dies, aging (wear-out) type probability distribution models (normal, lognormal, Weibull, and minimum extreme value) are well postulated. Such failures correspond to increasing failure rates. Each distribution reflects a slightly different failure rate model. Detailed description of curve fitting strategies in the reliability domain can be found in Sheikh et al. (2004).

Weibull distribution is widely used in reliability analyses for describing the distribution of time to failure (machines and components) and of strength (materials). It is particularly suited for situations where a weakest-link or the largest of many competing flaws is

responsible for failure. Minimum extreme value (EV) distribution is closely related to the Weibull distribution and is thus often used for similar purposes, such as representation of failure times. Lognormal and normal distributions are also closely related to each other, in the same way as Weibull and minimum EV distributions are. The mathematical ease to work with a Weibull model when deciding optimal die replacement strategies or determining number of dies needed for a given production run makes it more attractive.

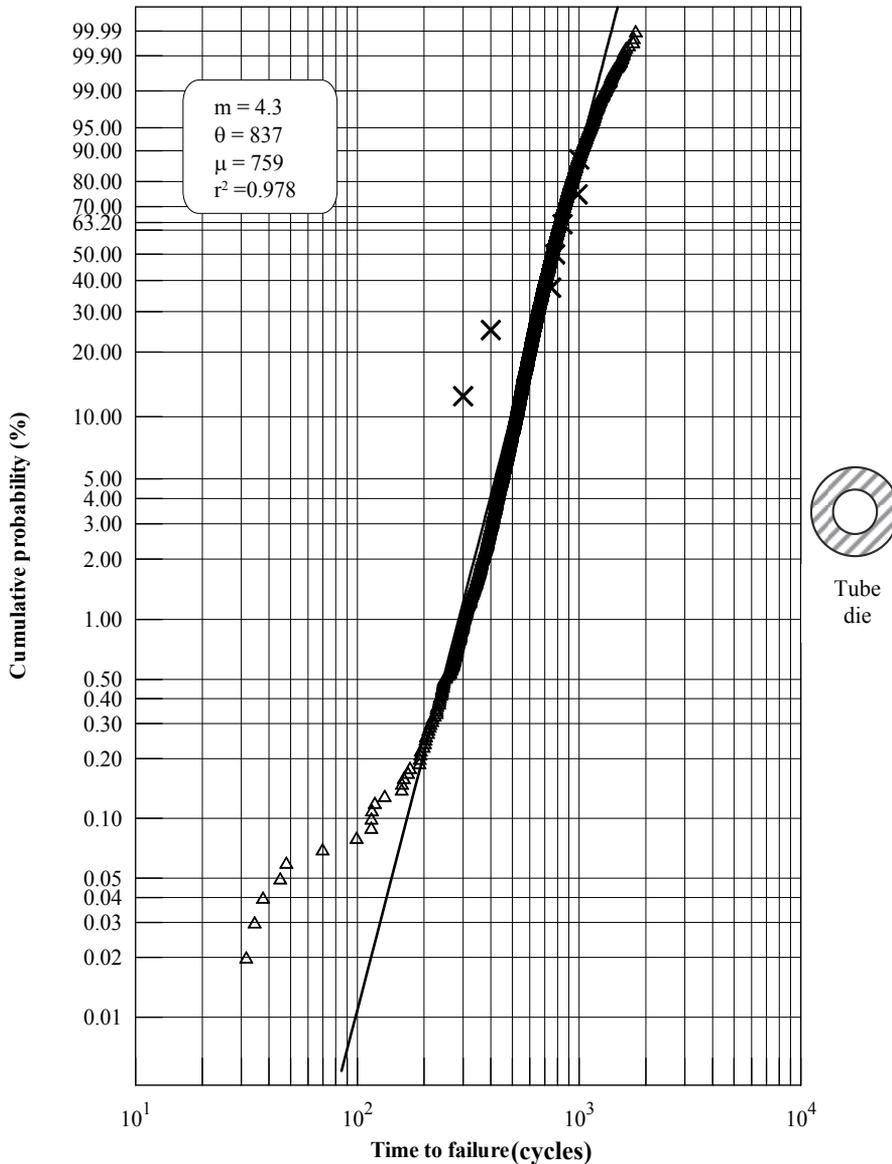


Fig. 6. Weibull model of fracture-failure simulation for the tube die

Once the simulated predictions for die failure by fracture are obtained, curve fitting is done to obtain standard probability distributions used in reliability studies (normal, lognormal,

Weibull, and minimum extreme value). Weibull distribution, shown in Fig. 6, gives the best overall goodness of fit (correlation coefficient of  $r^2 = 97.8\%$ ), although lognormal or even normal model cannot be ignored completely. Minimum extreme value distribution does not appear to be a good representation of die reliability. As evidenced by the graph, the Weibull line provides a good fit not only for the simulated data but also for the actual die life data (represented by  $x$  on the figure). Shape parameter for the Weibull line, representing the scatter in die life, came out to be  $m = 4.3$ . Value of the scale parameter  $\theta$  was 837 billets (cycles). Average die life is linked with the scale parameter and the scatter parameter by the equation

$$MTTF = \theta \Gamma(1 + 1/m), \quad (19)$$

where  $\Gamma(\ )$  is the gamma function. Thus, based on the simulation results,  $MTTF = 838 \Gamma(1 + 1/4.3) = 759$  cycles, whereas the actual  $MTTF$  due to fracture failures is 722 cycles.

#### 4. Wear failure

Saha (1998; 2000) studied the effect of extrusion speed and billet length on wear at the die (cap) and the mandrel shown in Fig. 7. A thin walled square tube of Al-6063 was extruded through heat treated and nitrided H13 steel die. Average billet temperature was  $460^\circ\text{C}$ . Higher extrusion speeds are expected to increase wear due to abrasion. Longer billet lengths generate larger friction surfaces, the increased friction promoting wear. Effect of ram speed and billet length on wear at die land was therefore studied.

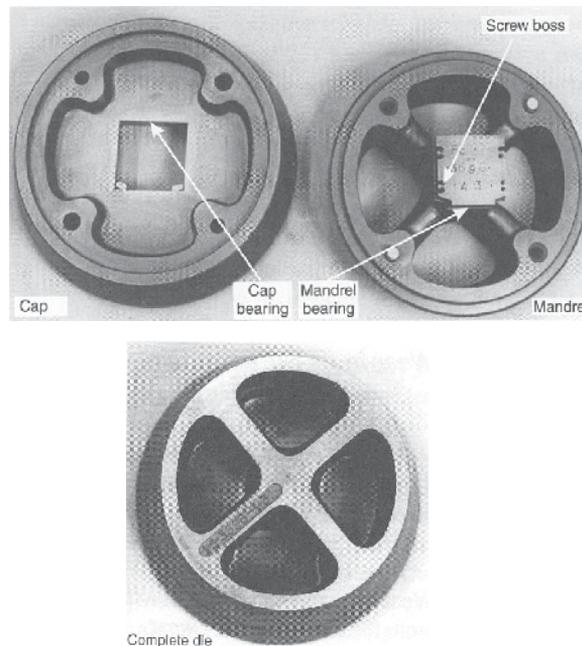


Fig. 7. Bearing surface on the die (cap) and mandrel, and the assembled die set used for wear experiments in hot aluminum extrusion (Saha, 2000)

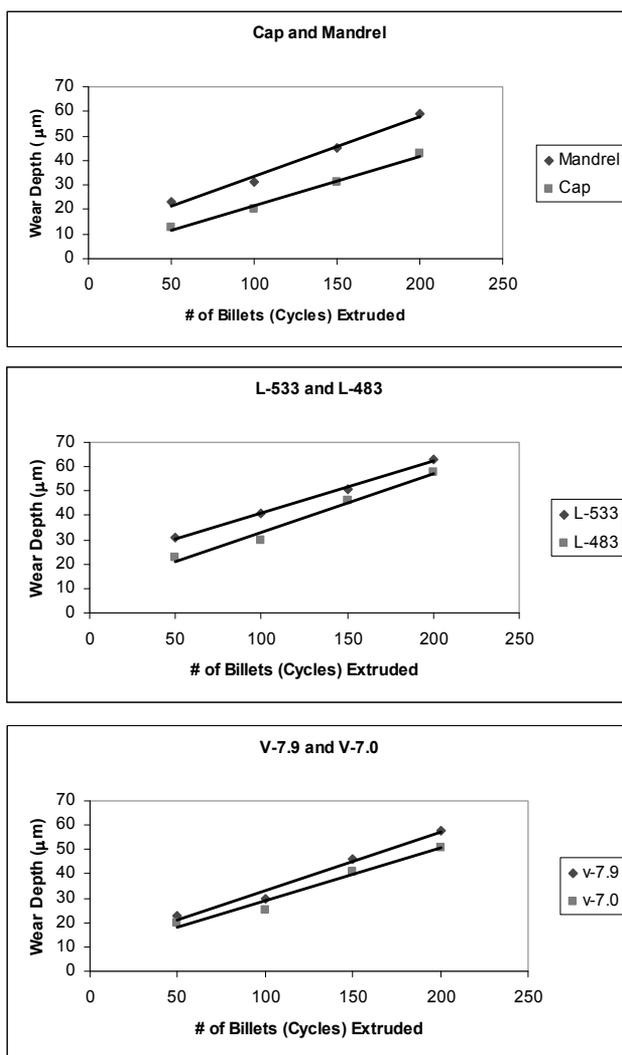


Fig. 8. Increasing wear depth at mandrel and cap of a hollow die for varying billet length and ram speed (billet temperature is 460°C); variation suggests linear behavior; data source Saha (2000)

Wear depth from these experiments (Saha, 2000) have been replotted in Fig. 8. The plots show that wear progresses almost linearly as more billets are extruded. A simple die wear model would thus be

$$W = at + c, \quad (20)$$

where  $W$  is the wear depth in microns ( $\mu\text{m}$ ),  $t$  is the number of extrusion cycles (billets extruded), and  $a$  and  $c$  are the slope and intercept of the straight line representing the progressive wear behavior. As a new die is perfectly smooth, with no wear at the bearing surface, we have  $W = 0$  at  $t = 0$ . The intercept would thus be zero. The model would then simplify to

$$W_f = at_f \quad (21)$$

$W_f$  signifies the limiting amount of wear leading to die failure, and  $t_f$  is the number of billets (cycles) extruded before reaching  $W_f$ .

#### 4.1 Parameter determination

Actual die failure data (due to wear at the die land) has been collected for the box profile described earlier. Die failure data (number of billets extruded before die was rejected due to excessive wear) of 21 initially-identical dies are given in Table 3. The average die life (MTTF) was 508 extrusion cycles. Based on the manufacturing tolerance for minimum wall thickness, it was determined that a failure wear of  $W_f = 75 \mu\text{m}$  on the bearing surface of the die or the mandrel would lead to rejection. This means that each time a die failure occurred, wear had reached this limiting value. Slope ( $a$ ) of the wear line was evaluated for each die rejection from the cycles (number of billets) to failure and limiting wear data. Mean and standard deviation values of this parameter ( $a$ ) came out to be 0.169 mm and 0.035. As demonstrated by Fig. 9, wear failure data for this box die are well represented by the Weibull distribution.

Failure #	Cycles to Failure	Failure #	Cycles to Failure
1	392	12	611
2	642	13	295
3	598	14	341
4	713	15	525
5	576	16	402
6	335	17	356
7	340	18	282
8	548	19	227
9	480	20	658
10	725	21	1050
11	559		

Table 3. Various instances of number of billets (cycles) to failure for the box-die studied, indicating when surface wear reaches a critical value of  $W_f = 75$  microns

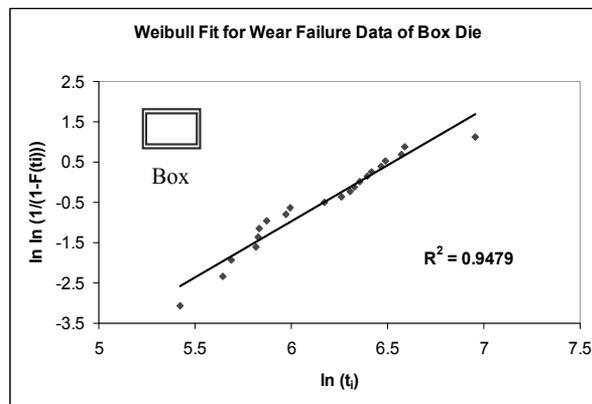


Fig. 9. Wear failure data for the box die are well represented by the Weibull distribution ( $r^2 = 94.8\%$ )

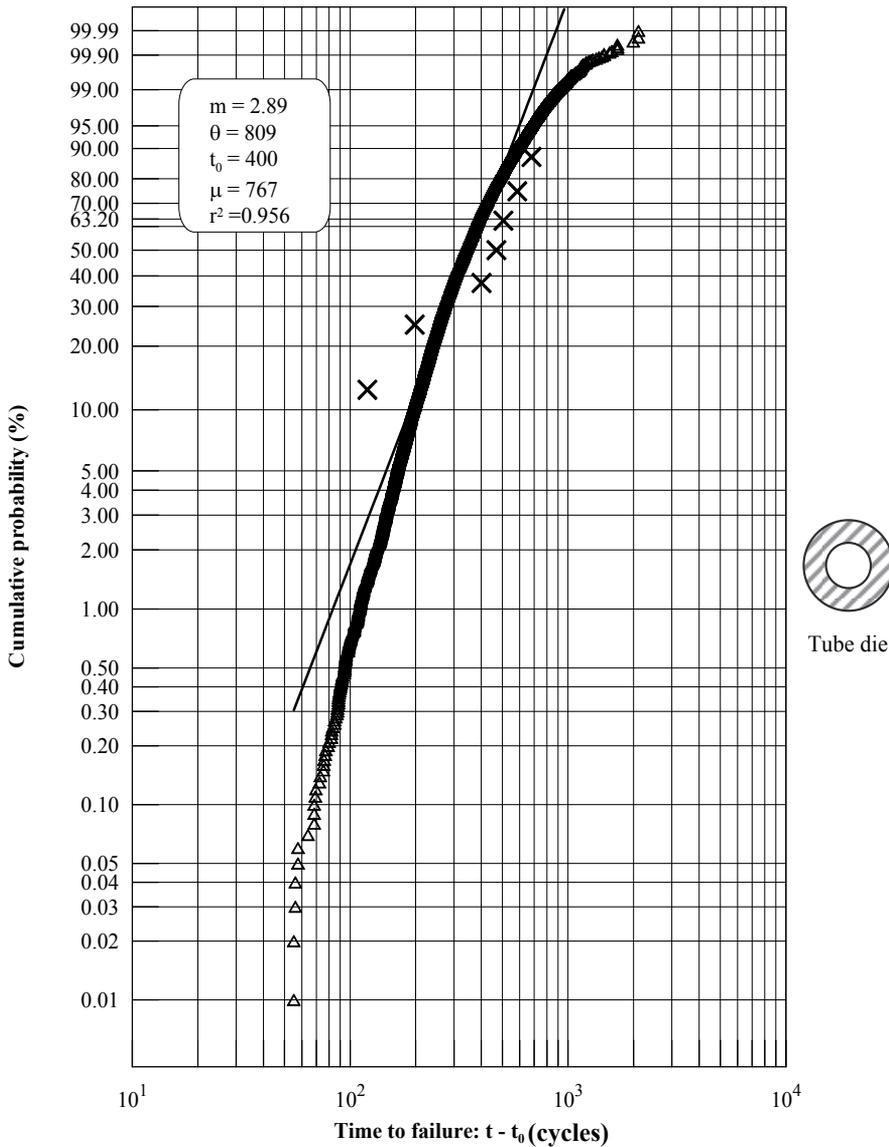


Fig. 10. Plot showing simulated wear life data and the fitted 3-parameter Weibull line for the tube die

#### 4.2 Monte Carlo simulation — wear

Failure data of the box die were used only to generate the wear model described above. Monte Carlo simulation was then carried out for the same hollow die (simple tube) whose fracture simulation was presented before. It was assumed that the distribution (mean and standard deviation) of the slope  $m$  characterizing wear behavior remains the same for all H13 steel dies used in hot aluminum extrusion. 10,000 standard normal random numbers were generated as before, and were transformed to random values of the slope  $m$ , based on its  $\mu$  and  $\sigma$  values determined above. A limiting wear value of 125 microns was used,

determined from the dimensional tolerances for the tube die. Wear life of the die was then evaluated for each simulated instance of  $m$ , using equation (21). Compared to an actual MTTF value of 722 billets, the simulated life (based on wear failures only) came out to be 767 cycles. We know from the actual failure history of this die that wear was not a dominant failure mode. This fact is confirmed by the Weibull fit to the simulated data (Fig. 10), yielding a lower correlation coefficient of 95.6% (as against  $r^2 = 97.8\%$  for fracture failure).

## 5. Combined fracture and wear

As explained earlier, various failure mechanisms are operating simultaneously on the die during its operative life. Fracture and wear are thus competing against each other. Final die failure takes place either when a preexisting crack (of size  $a_0$ ) reaches the critical crack size ( $a_c$ ), or when wear on the die land reaches the limiting value ( $W_f$ ). It is assumed here that the two failure modes progress independently of each other, until one of them becomes dominant and takes over. However, in reality, the failure modes may be inter-dependent in a rather complicated manner.

The number of cycles to failure under the competing fracture-wear mode was determined from the relation

$$t = \min(t_F, t_W), \quad (22)$$

where  $t_F$  is the simulated fracture life,  $t_W$  is the wear life, and  $t$  is the final predicted life, representing the failure (fracture or wear) that occurs earlier. The MTTF of this simulated fracture-wear die life was 717 billets. In comparison with the actual average life of 722 billets, this is a very close prediction. Weibull distribution was once again found to be the best fitted to the simulated die life data, with a coefficient of correlation value of  $r^2 = 98.5\%$ , as shown in Fig. 11. This combined fracture-wear model is obviously the best representation of die failure, yielding a shape parameter  $m = 4.99$  and a scale parameter  $\theta = 786$  billets.

## 6. Conclusions

This chapter presents a simulation strategy for prediction of service life of metalworking dies and tools that undergo fracture and wear. A concise review of fracture and wear failures of hot-work extrusion dies, and of the Monte Carlo simulation strategy has been presented. Failure mechanisms under fatigue and wear are briefly explained, using basic fracture mechanics and tribology principles. Mathematical models for the two failure types have been explained. Die life and related material and geometrical parameters have been treated as random variables, the approach being closer to reality as compared with the deterministic models usually employed. Monte Carlo simulation has been carried out to predict the life of an extrusion die, Paris law providing the physical-mathematical model for estimation of fatigue life of the die in terms of number of cycles to failure (number of billets extruded). Correlation models developed earlier by the authors have been utilized to estimate the value of fracture toughness ( $K_{IC}$ ) of H13 hot work die steel under typical tempering and operating temperature conditions. Failure history of an actual simple hollow die from the aluminum extrusion industry has been used as a case study.

The resulting (simulated) die life observations are adequately represented by a Weibull probability model with shape parameter  $m = 4.3$  and scale parameter  $\theta = 837$  cycles (billets), an average die life of 759 cycles (as against the actual MTTF of 722 cycles), and a correlation coefficient of  $r^2 = 97.8\%$ . A failure model based on surface wear has also been developed,

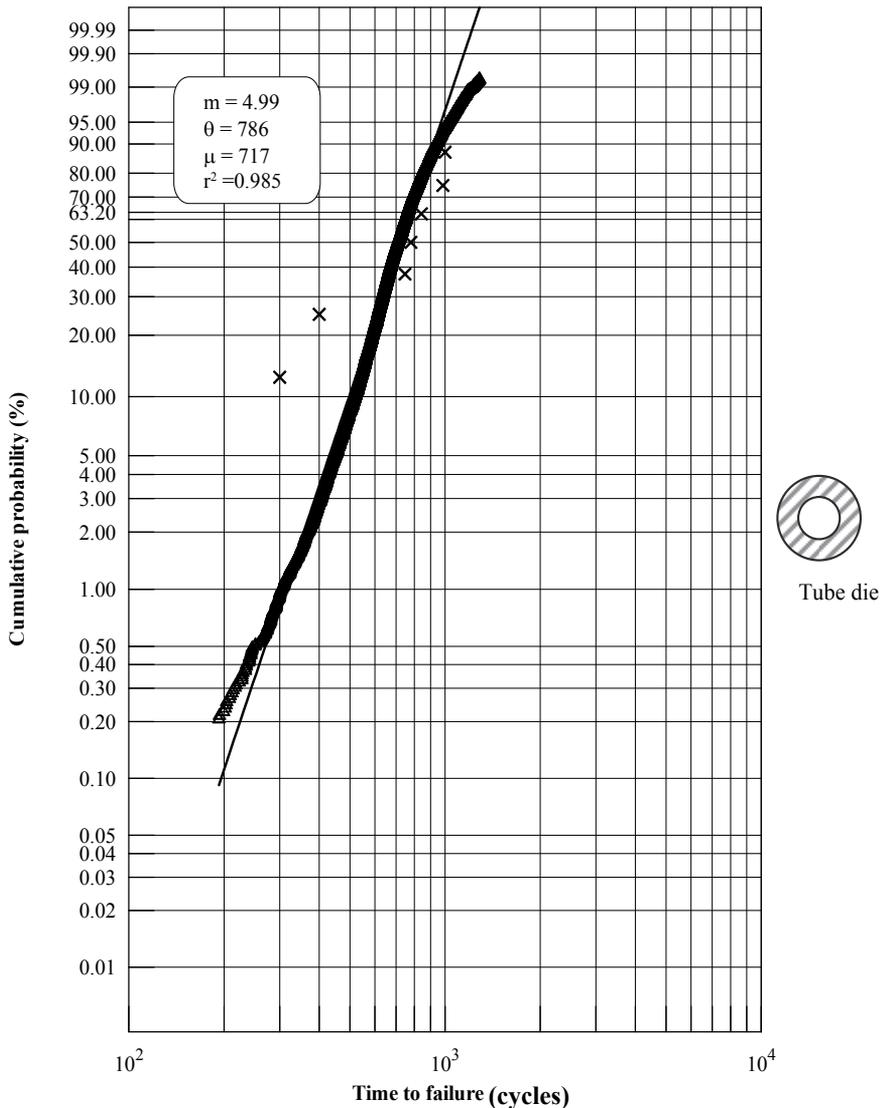


Fig. 11. Plot of Weibull model for simulated combined life data (fracture-wear competing mode) for the tube die

using actual wear data of a hollow box die. Using this linear wear-failure model, Monte Carlo simulation has been carried out to predict the die life of the tube die. The simulation yielded a Weibull fit with  $m = 2.89$ ,  $\theta = 809$  cycles, MTF = 767 cycles, and  $r^2 = 95.6\%$ . This lower goodness of fit for wear matches well with failure records, as the tube die did not fail frequently due to wear in actual practice. The combined fracture-wear failure model yielded the best simulation, with  $m = 4.99$ ,  $\theta = 786$  cycles, MTF = 717 cycles, and  $r^2 = 98.5\%$ .

The strategy outlined here can be easily adapted to forecast the fracture, wear, or combined fracture-wear life of cold-work or hot-work metal-forming dies (especially extrusion) that are made from tool steels, are subjected to different heat treatment routines, and are used at different operating temperatures.  $K_{IC}$ -CVN correlations developed by the authors (Qamar et

al.; 2006) can be used to find the values of fracture toughness of heat-treated tool steels. The work can be extended to cover more complex profiles, extrusion pressure for a specific profile being estimated using its shape complexity (Qamar et al., 2004). Expected service life can thus be predicted for a die of a given profile even before the die is used. The industry can employ the life prediction scheme to develop *optimum die replacement strategies*. This can contribute to serious reduction in warehousing costs, and can prevent unnecessary downtime due to unavailability of a certain die profile, thereby increasing productivity and client satisfaction. Analysis of die-life simulation results can also help improve die design against fracture and wear failures.

### 6.1 Future work; special note

A simulation is as good as the underlying model. Two basic simplifications are made in this study: using the model of an edge crack in a flat plate whereas the actual die is a hollow cylinder, and disregarding the correction factor for finite width.  $K_{IC}$ -CVN correlations used for estimation of hot fracture toughness may need more experimental work and refinement. Paris equation parameters  $C$  and  $m$  used here are for a generic hot-work tool steel; actual values of these constants may be needed for H13 tool steel (not available till now). With these refinements, the model can be initially used for more accurate simulation of failures of dies of simple geometry. The simulation strategy could then be used to forecast failures of more complex dies by employing the complexity-pressure models developed earlier by the authors (Qamar et al., 2004). A relatively sound replacement strategy for dies and related tooling can thus be formulated.

## 7. References

- Anderson, T.L. (2005) *Fracture Mechanics Fundamentals and Applications*, 3rd edition, Taylor and Francis, ISBN-13 978-0-8493-1656-2, Boca Raton, Louisiana
- Arif, A.F.M.; Sheikh, A.K.; Qamar, S.Z.; Al-Fuhaid, K.M. (2003) A study of die failure mechanisms in aluminum extrusion, *Journal of Materials Processing Technology*, Vol. 134, No. 3, p 318-328
- Barsom, J.M.; Rolfe, S.T. (1970) Correlations between  $K_{IC}$  and Charpy V-notch test results in the transition-temperature range, *Impact Testing of Metals*, ASTM STP 466, American Society for Testing and Materials, Philadelphia, p 281-302
- Bauser, M.; Sauer, G.; Siegert, K. (2006) *Extrusion*, 2<sup>nd</sup> edition, ASM International, ISBN-13 : 978-0-87170-837-3, Materials Park, Ohio
- Björk T, Westergard R, Hogmark S (2001) Wear of surface treated dies for aluminum extrusion – a case study, *Wear*, Vol. 249, p 316-323
- Budynas, R.G.; Nisbett J.K. (2008) *Shigley's Mechanical Engineering Design*, 8th edition, McGraw-Hill, ISBN 978-007-125763-3, Singapore
- Callister, W.D. (2006) *Materials Science and Engineering: An Introduction*, 6th edition, John Wiley, ISBN 8126508132, New York
- Cosenza, C.; Fratini, L.; Pasta, A.; Micari, F. (2004) Damage and fracture study of cold extrusion dies, *Engineering Fracture Mechanics*, Vol. 71, p 1021-1033
- Gouveia, P.A.; Rodrigues, J.M.C.; Martins, P.A.F. (2000) Ductile fracture in metalworking: experimental and theoretical research, *Journal of Materials Processing Technology*, Vol. 101, No. 1-3 (April 2000), p 52-63
- Groover, P.G. (2010) *Fundamentals of Modern Manufacturing: Materials, Processes, and Systems*, John Wiley, ISBN-13 978-0470467008, New York

- Hambli, R.; Badie-Levet, D.; (2000) Damage and fracture simulation during the extrusion process, *Computer Methods in Applied Mechanics and Engineering*, Vol. 186, No. 1, p 109-120
- Laue, K. ; Stenger, H. (1981) *Extrusion: Processes, Machinery, Tooling*, American Society for Metals, ISBN 0-87170-094-8, Metals Park, Ohio
- Lee, G-A.; Im, Y-T.; (1999) Finite element investigation of the wear and elastic deformation of dies in metal forming, *Journal of Materials Processing Technology*, Vol 89-90 (May 1999), p 123-127
- Müller, K.B. (2002) Deposition of hard films on hot-working steel dies for aluminum, *Journal of Materials Processing Technology*, Vol. 130-131, p 432-437
- Nanninga, N.; White, C. (2009) The relationship between extrusion die line roughness and high cycle fatigue life of an AA6082 alloy, *International Journal of Fatigue*, Vol. 31, No. 7 (July 2009), p 1215-1224
- Pöhlandt, K.; Kuehl, R. (1989) *Materials Testing for the Metal Forming Industry*, Springer-Verlag, ISBN 0387506519, Berlin
- Qamar, S.Z.; Arif, A.F.M.; Sheikh, A.K. (2004) A new definition of shape complexity for metal extrusion, *Journal of Materials Processing Technology*, Vol. 155-156, No. 30 (November 2004), p 1734-1739
- Qamar, S.Z.; Sheikh, A.K.; Arif, A.F.M.; Pervez, T. (2006) Regression-based  $CVN-K_{IC}$  models for hot work tool steels, *Materials Science and Engineering A*, Vol. 430, No. 1-2 (August 2006), p 208-215
- Saha P K (1998) Thermodynamics and tribology in aluminum extrusion, *Wear*, Vol. 218, p 179-190
- Saha, P. K. (2000) *Aluminum Extrusion Technology*, ASM International, ISBN-13 978-0871706447 Materials Park, Ohio
- Sanford, R.J. (2003) *Principles of Fracture Mechanics*, Pearson Education, ISBN-13 978-0130929921, Upper Saddle River, New Jersey
- Sheikh AK, Arif AFM, Qamar SZ (2004) A probabilistic study of failures of solid and hollow dies in hot aluminum extrusion, *Journal of Materials Processing Technology*, Vol. 155-156, No. 30 (November 2004), p 1740-1748
- Sheppard, T. (1999) *Extrusion of Aluminum Alloys*, Kluwer Academic Publishers, ISBN 0 412 59070 0, Dordrecht
- So, H.; Chen, H.M.; Chen, L.W. (2008) Extrusion wear and transition of wear mechanisms of steel, *Wear*, Vol. 265, No. 7-8 (September 2008), p 1142-1148
- Sudhakar, K.V.; (2002) Micromechanics of fracture in extrusion die punch, *Engineering Failure Analysis*, Vol. 9, p 159-165
- Terčelj, M.; Smolej, A.; Fajfar, P.; Turk R. (2007) Laboratory assessment of wear on nitrided surfaces of dies for hot extrusion of aluminium, *Tribology International*, Vol. 40, No. 2 (February 2007), p 374-384
- Thedja W W, Muller B K, Ruppin D (1992) Tribological processes on the die land area during extrusion of Al6063," *Proceedings 5th International Aluminum Extrusion Technology Seminar*, Aluminum Association & Aluminum Extruders Council
- Tseronis, D. ; Sideris, I.F. ; Medrea, C.; Chicinas, I. (2008) Microscopic examination of the fracture surfaces of an H13 hot extrusion die due to failure at the initial usage stage, *Key Engineering Materials*, Vol. 367 (2008), p 177-184
- Yoh, E-G.; Kim, Y-I.; Lee, Y-S.; Park, H-J.; Na, K-H.; (2002) Integrated analysis for die design including brittle damage evolution, *Journal of Materials Processing Technology*, Vol. 130-131 (December 2002), p 647-652

# Loss of Load Expectation Assessment in Electricity Markets using Monte Carlo Simulation and Neuro-Fuzzy Systems

H. Haroonabadi

Islamic Azad University (IAU)-Islamshahr Branch  
Iran

## 1. Introduction

The power systems main emphasis is to provide a reliable and economic supply of electrical energy to the customers (Billinton & Allan, 1996). A real power system is complex, highly integrated and almost very large. It can be divided into appropriate subsystems in order to be analyzed separately (Billinton & Allan, 1996). This research deals with generation reliability assessment in power pool markets, and transmission and distribution systems are considered reliable (Hierarchical Levels-I, HL-I) as shown in Fig. 1.

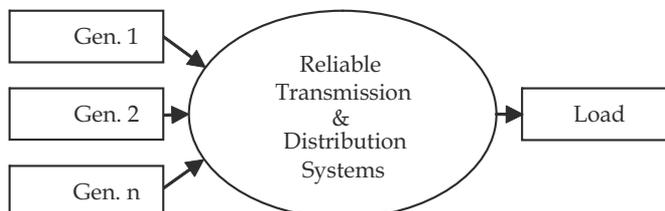


Fig. 1. Power pool market schematic for generation reliability assessment

Most of the methods used for generation reliability evaluation are based on the “loss of load or energy” approach. One of the suitable indices that describes generation reliability level is “Loss of Load Expectation” (*LOLE*), that is the time in which load is more than the available generation capacity.

Generally, the reliability indices of a system can be evaluated using one of the following two basic approaches (Billinton & Allan, 1992):

- Analytical techniques
- Stochastic simulation

Simulation techniques estimate the reliability indices by simulating the actual process and random behavior of the system. Since power markets and generators’ forced outages have stochastic behavior, Monte Carlo Simulation (MCS), as one of the most powerful methods for statistical analysis of stochastic problems, is used for reliability assessment in this research.

Generation reliability depends absolutely on the generating units specifications. The main function in traditional structure for Unit Commitment (UC) of the generators is to minimize generation costs. Since the beginning of the 21<sup>st</sup> century, many countries have been trying to deregulate their power systems and create power markets (Salvaderi, 2000), (Mountford & Austria, 1999), (Draper, 1998), (Puttgen et al, 2001), (Mc Clanahan,2002). In the power markets, the main function of players is their own profit maximization, which severely depends on the type of the market. As a result, generation reliability assessment depends on market type and its characteristics.

Generally, economists divide the markets into four groups, varying between perfect competition market and monopoly market (Pindyck & Rubinfeld, 1995). This study deals with the evaluation of generation reliability in different kinds of power pool markets based on the market concentration. Let's review some of the papers proposed till now.

An optimization technique is proposed in (Wang et al, 2009) to determine load shedding and generation re-dispatch for each contingency state in the reliability evaluation of restructured power systems with the Poolco market structure. The problem is formulated using the optimal power flow (OPF) technique. The objective of the problem is to minimize the total system cost, which includes generation, reserve and interruption costs, subject to market and network constraints.

Reference (Jaeseok et al, 2001) has used "Effective Load Duration Curve" (ELDC) for evaluation of "Loss of Load Expectation" (LOLE) and "Expected Energy Not Served" (EENS) as reliability indices.

Reference (Wang & Billinton, 2001) has presented some reliability models for different players in a power system, where generation system is represented by an equivalent multi-state generation provider (EMGP). The reliability parameters of each EMGP are shown by an available capacity probability table (ACPT), which is determined using conventional techniques. Then, the equivalent reliability parameters for each state (including state probability, frequency of encountering the state and the equivalent available generation capacity) are determined.

Reference (Haroonabadi & Haghifam, 2009) compares generation reliability in various economic markets: Perfect Competition, Oligopoly and Monopoly power pool markets. Also, due to the stochastic behavior of power market and generators' forced outages, Monte Carlo Simulation is used for reliability evaluation.

In researches dealing with power marketing and restructuring, market behavior and its economic effects on the power system should be considered. Therefore, this research considers power pool market fundamentals and deals with generation reliability assessment in power pool market using MCS and an intelligent system. Also, sensitivity of reliability index to different reserve margins and times will be evaluated. In Section-2, the fundamentals of power pool market will be discussed. In Section-3, the algorithm for generation reliability assessment in power pool market will be proposed, and finally in Section-4, the case study results will be presented and discussed.

## 2. Power pool markets fundamentals

Market demand curve has negative gradient, and the amount of demand decrease is explained by "price elasticity of demand". This index is small for short terms, and big for

long terms; because in longer terms, customers can better adjust their load relative to price (IEA, 2003). Demand function, generally, is described as  $P=a-b.Q$ . Therefore, price elasticity of demand is explained as:

$$E_d = \left| \frac{dQ}{dP} \right| = \frac{1}{b} \tag{1}$$

Let's suppose forecasted load by dispatching center is an independent power from price that equals to  $Q_n$ . Therefore, demand function can be obtained as:

$$P = a - b.Q = b.Q_n - b.Q = \frac{Q_n}{E_d} - \frac{Q}{E_d} \tag{2}$$

Typically, as shown in Fig. 2, price elasticity in power markets is 0.1-0.2 for the next 2-3 years and 0.3-0.7 for the next 10-20 years (IEA, 2003).

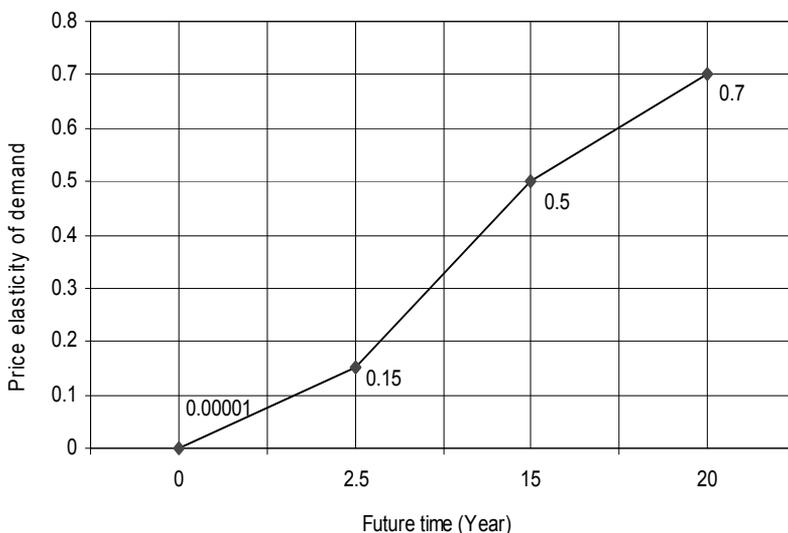


Fig. 2. Price elasticity of demand for various times

Offer curve of a company, which participates in a market without any market power is part of the marginal cost curve that is more than minimum average variable cost (Pindyck & Rubinfeld, 1995). Also, total offer curve of all companies is obtained from horizontal sum of each company's offer curve. This curve is a merit order function. In economics, if sale price in a market becomes less than minimum average variable cost, the company will stop production; because the company will not be able to cover not only the fix cost but even the variable cost (Pindyck & Rubinfeld, 1995). Due to the changing efficiency and heat rate of power plants, marginal cost is less than average variable cost. Therefore, in power plants, average variable cost replaces marginal cost in economic studies (Borenstein, 1999).

In a perfect competition market, equilibrium price and equilibrium amount are obtained from the intersection of total offer curve and demand curve. On the other hand, in a

monopoly market, the monopolist considers the production level, which maximizes his profit. It is proved that the monopolist considers the level of production in which marginal cost of each firm (and total marginal cost of all firms) equals to the marginal revenue of the monopolist (Pindyck & Rubinfeld, 1995):

$$MC_1 = MC_2 = \dots = MC = MR \quad (3)$$

Where:

$$MR = a - 2.b.Q = b.Q_n - 2.b.Q = \frac{Q_n}{E_d} - \frac{2.Q}{E_d} \quad (4)$$

Comparison of (2) and (4) shows that if there is no market power, offer curve of industry for each market (from perfect competition market to monopoly market) will equal marginal cost; but negative gradient of demand exponent curve (DE) varies between  $b$  (for demand function in perfect competition market) and  $2b$  (for marginal revenue in monopoly market). Therefore, generally, demand exponent curve can be expressed as:

$$DE = a - K.b.Q = \frac{Q_n}{E_d} - \frac{K.Q}{E_d} \quad (5)$$

Where,  $K$  varies between 1 and 2.

Fig. 3 shows the typical total offer and demand exponent curves.

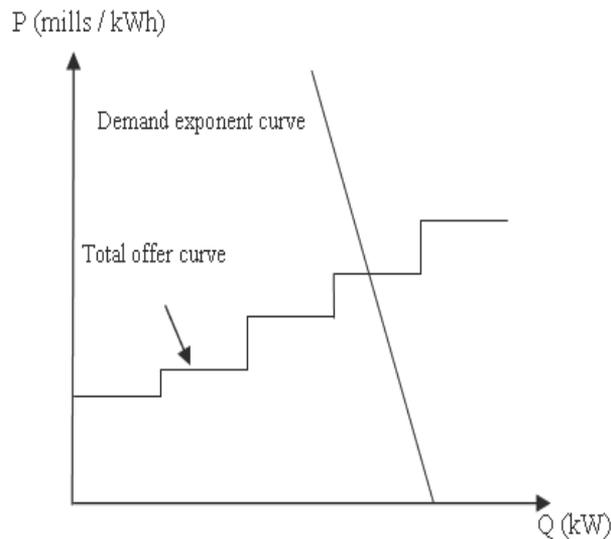


Fig. 3. Typical total offer and demand exponent curves

### 3. Proposed method for generation evaluation in power markets

In power markets, Hirschman-Herfindahl Index ( $HHI$ ), which is obtained from (6), is used for market concentration measurement (IEA, 2003):

$$HHI = \sum_M q_i^2 \tag{6}$$

If market shares are measured in percentages, *HHI* will vary between 0 (an atomistic market) and 10000 (monopoly market). According to a usual grouping, the US merger guidelines stipulate an assumption that markets with a *HHI* below 1000 is unconcentrated, a *HHI* between 1000 and 1800 is moderately concentrated, and a *HHI* above 1800 is highly concentrated (FTC, 1992).

As mentioned before, according to the type of market and *HHI* values, negative gradient of demand exponent curve varies between *b* and *2b*. Therefore, for modeling the market, a fuzzy number is proposed in this study to estimate the gradient coefficient of demand exponent curve (*K*) based on the *HHI* values. Membership functions of unconcentrated, moderately concentrated and highly concentrated markets' fuzzy sets and the equation to estimate gradient coefficient are shown in Fig. 4 and (7), respectively.

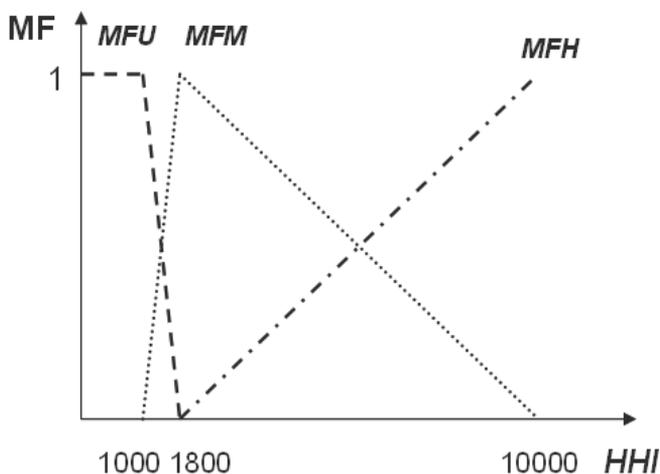


Fig. 4. Membership functions of unconcentrated, moderately concentrated and highly concentrated markets' fuzzy sets

$$K = (MFU + 1.5 \times MFM + 2 \times MFH) \tag{7}$$

As Fig. 4 and (7) show, while the proposed coefficient (*K*) covers all kinds of markets with different concentration degrees, the changes of these degrees are not sudden, rather they are gradual and continuous. Also, the proposed method and fuzzy logic are valid for all power pool markets.

Generation reliability of a power system depends on many parameters, especially on reserve margin, which is defined as (IEA, 2002):

$$RM\% = \frac{Installed\ Capacity - Peak\ Demand}{Peak\ Demand} \times 100 \tag{8}$$

The algorithm of generation reliability assessment in power pool markets using Monte Carlo simulation and proposed fuzzy logic is as follows (Fig. 5):

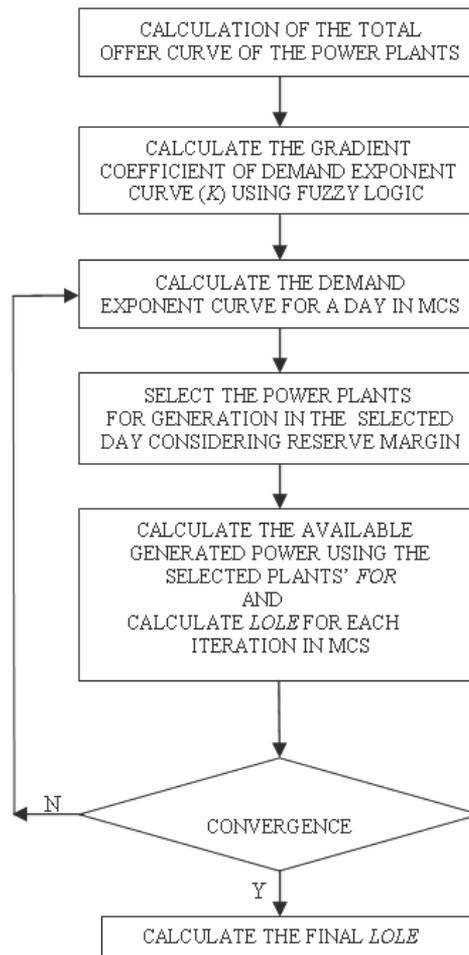


Fig. 5. Flow chart of HLI reliability assessment in power markets using MCS

$HHI$  is obtained based on characteristic of the market. The gradient coefficient of demand exponent curve ( $K$ ) is calculated using Fig. 4 and (7).

1. Calculation of the total offer curve of power plants.
2. Select a random day and its load ( $Q_n$ ), and calculate demand exponent curve using (5).
3. The power plants, selected for generation in the selected day, are determined from the intersection of the power plants' total offer curve and demand exponent curve with regards to the reserve margin.
4. For each selected power plant in the previous step, a random number between 0-1 is generated. If the generated number is more than the power plant's Forced Outage Rate ( $FOR$ ), the power plant is considered as available in the mentioned iteration; otherwise it encounters forced outage and thus can not generate power. This process is performed for all power plants using an independent random number generated for each plant. Finally, sum of the available power plants' generation capacities is calculated. If the sum becomes less than the intersection of power plants' total offer curve and demand exponent curve, we will have interruption in the iteration, and therefore,  $LOLE$  will

increase one unit; otherwise, we will go to the next iteration. The algorithm of available generated power and *LOLE* calculations for each iteration in MCS is shown in Fig. 6.

5. The steps 3 to 5 are repeated for calculation of final *LOLE*.

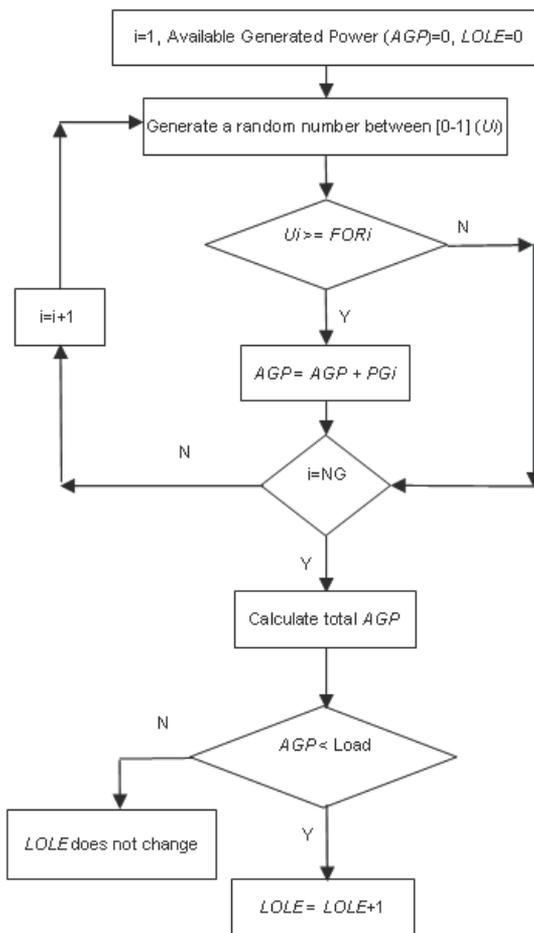


Fig. 6. Algorithm of available generated power and *LOLE* calculations for each iteration using MCS

Now, to create a unique structure, a four-layer perceptron neural network (N.N.) is used for reliability evaluation. The number of the neurons in each layer is 20, 15, 12 and 1, respectively (Fig. 7). All the neurons in the first, third and last layers have POSLIN transfer function, and the second layer has TANSIG transfer function. Inputs of the neural network include:

- Gradient coefficient of demand exponent curve (*K*)
- Simulated future time (*FT*)
- Reserve margin (*RM*)

Also, neural network's output is *LOLE* index.

Parts of the MCS results, obtained from the mentioned algorithm, are used for neural network training.

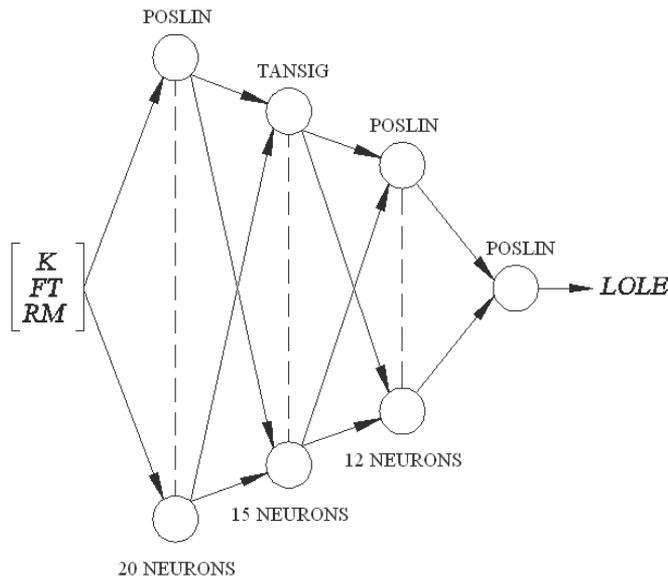


Fig. 7. Proposed N.N. for HLI reliability evaluation

#### 4. Numerical studies

IEEE - Reliability Test System (IEEE-RTS) is used for case studies. The required data for IEEE-RTS can be found in (Reliability Test System..., 1979). The following assumptions are used in various case studies:

1. All case studies are simulated for the second half of the year, based on the daily peak load of the mentioned test system.
2. All simulations are done with 5000 iterations.
3. Neural network is trained with TRAINLM method in MATLAB 7.0 software with 150 epochs. In this research, the neural network reached 0.2 Mean Square Error (MSE) after training.
4. Each case study is simulated for two different times (present time and the 2<sup>nd</sup> next year) and two different reserve margins (0%, 9%).
5. Annual growth rates of the power plants' generation capacity and consumed load are considered as 3.4% and 3.34%, respectively.
6. Annual growth rates of oil and coal costs are considered as 4% and 1%, respectively. Nuclear fuel cost (including uranium, enrichment and fabrication) is considered as a fixed rate. Also, annual growth rate of variable Operating and Maintenance (O&M) cost is considered as 1%.

In the first case study, each power plant is assumed as an independent company. Therefore, *HHI* equals 634, and the market is unconcentrated. Using Fig. 4 and (7), *K* is calculated as 1, as shown in Fig. 8. Based on this assumption and using MCS algorithm and the proposed neural network, *LOLE* values are obtained versus different times and reserve margins as shown in Fig. 9 and Fig. 10, respectively.

The error between the *LOLE* values obtained from MCS and neural network in the first study is 0.4%.

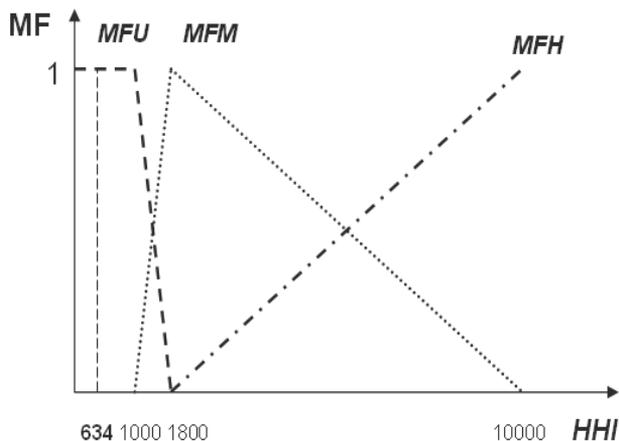


Fig. 8. The gradient calculation of demand exponent curve using membership functions for the first study

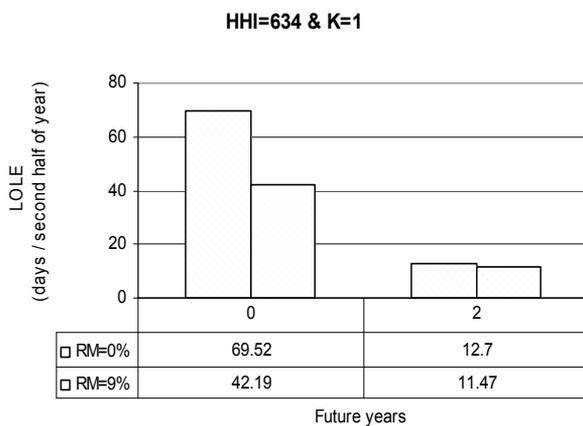


Fig. 9. LOLE values for the first study using MCS

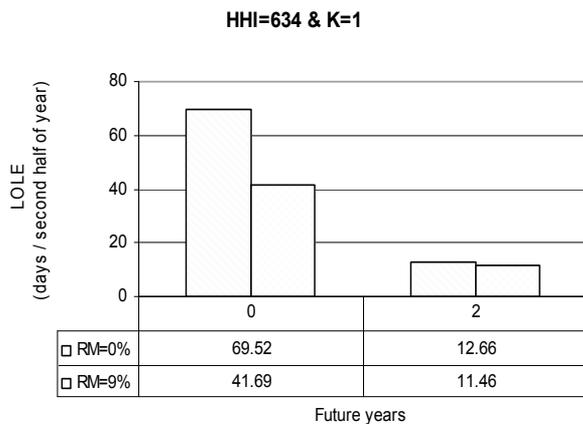


Fig. 10. LOLE values for the first study using N.N.

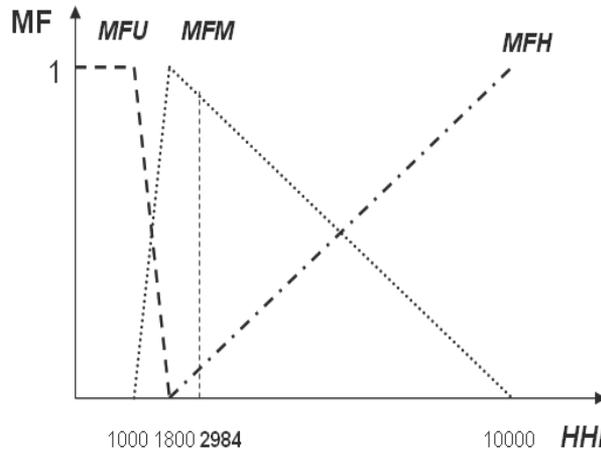


Fig. 11. The gradient calculation of demand exponent curve using membership functions for the second study

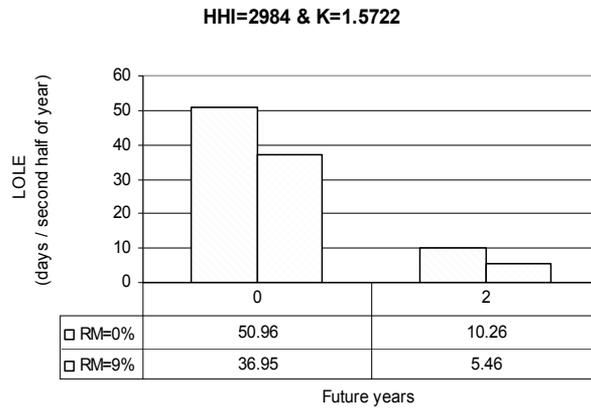


Fig. 12. LOLE values for the second study using MCS

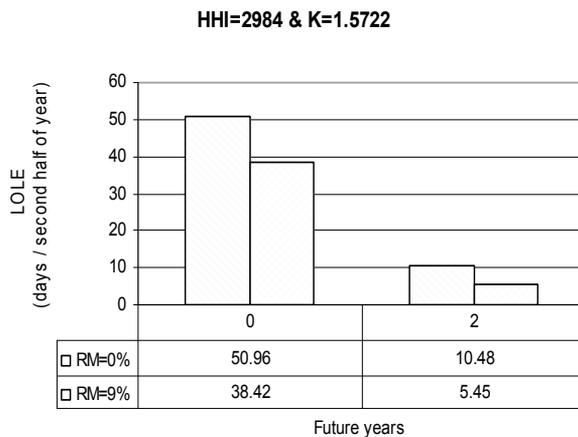


Fig. 13. LOLE values for the second study using N.N.

In the second study, all the power plants based on their types (including oil, coal, nuclear and water plants) are classified. Therefore,  $HHI$  equals 2984, and  $K$  is calculated as 1.5722 (Fig. 11). Based on this assumption and using MCS algorithm and the proposed neural network,  $LOLE$  values are obtained versus different MCS times and reserve margins as shown in Fig. 12 and Fig. 13, respectively.

The error between the  $LOLE$  values obtained from MCS and neural network in the second study is 1.64%.

In the third study, all fossil power plants (including oil and coal power plants) are classified in one company, and other power plants are as in the second case study. Therefore, the types of power plants are fossil, nuclear and water. As a result,  $HHI$  equals 5290, and  $K$  is calculated as 1.7128 (Fig. 14). Based on this assumption and using MCS algorithm and the proposed neural network,  $LOLE$  values are obtained versus different times and reserve margins as shown in Fig. 15 and Fig. 16, respectively.

The error between the  $LOLE$  values obtained from MCS and neural network in the third study is 1.53%.

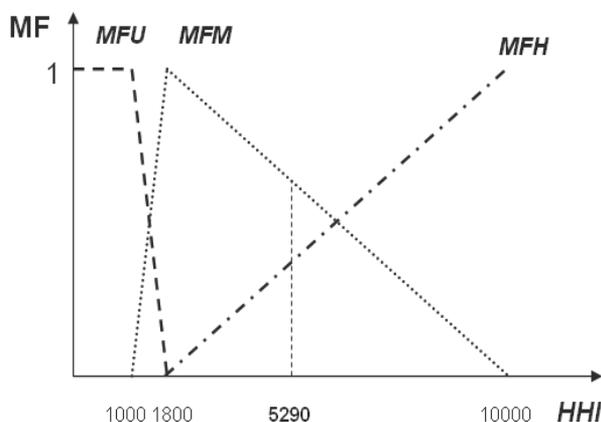


Fig. 14. The gradient calculation of demand exponent curve using membership functions for the third study

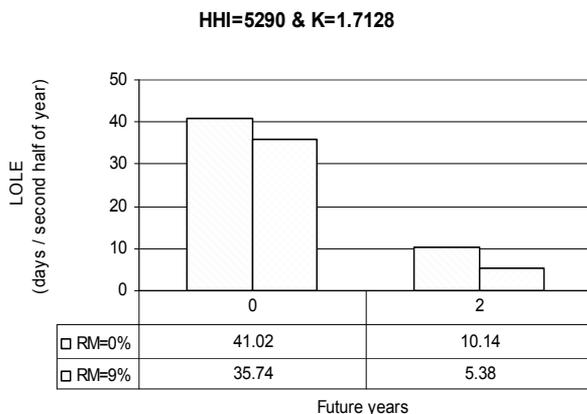


Fig. 15.  $LOLE$  values for the third study using MCS

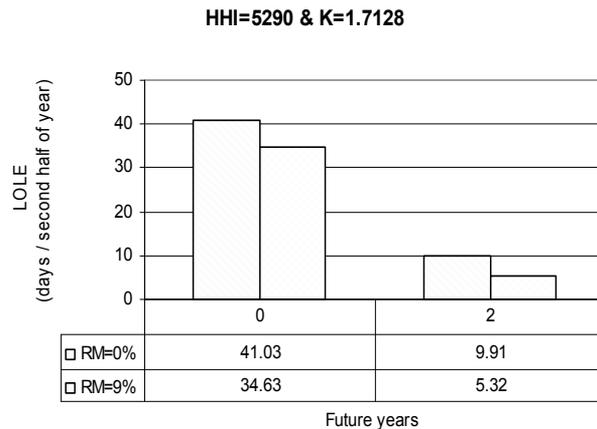


Fig. 16. *LOLE* values for the third study using N.N.

In the fourth study, it is assumed that all power plants belong to a monopolist, and the market is fully concentrated and monopoly. Therefore, *HHI* equals 10000, and *K* is calculated as 2 (Fig. 17). Based on this assumption and using MCS algorithm and the proposed neural network, *LOLE* values are obtained versus different times and reserve margins as shown in Fig. 18 and Fig. 19, respectively.

The error between the *LOLE* values obtained from MCS and neural network in the fourth study is 0.5%.

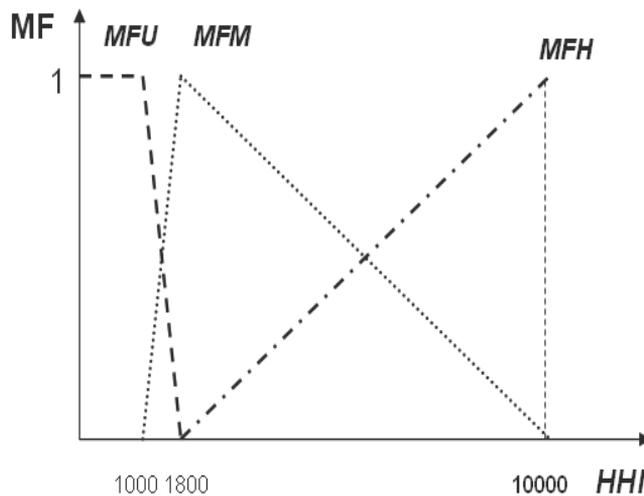


Fig. 17. The gradient calculation of demand exponent curve using membership functions for the fourth study

As it is shown in all case studies, *LOLE* values in the neural network method are very similar to MCS values. Evidently, the neural network's specifications depend on the power system's characteristics, and the proposed neural network is valid for the mentioned power system. Therefore, neural network's specifications may be changed in another power system based on the power system parameters.

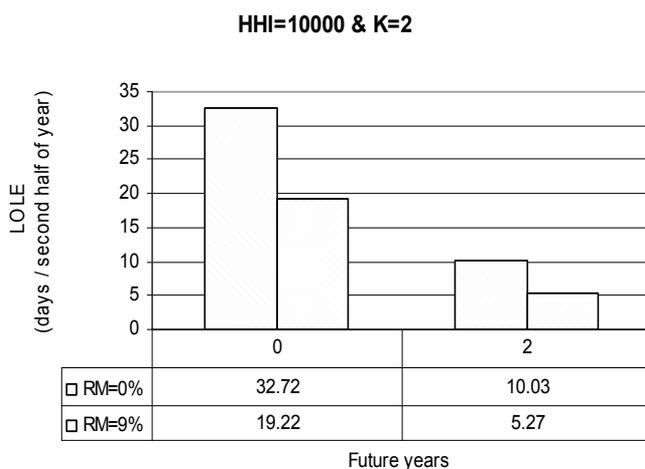


Fig. 18. *LOLE* values for the fourth study using MCS

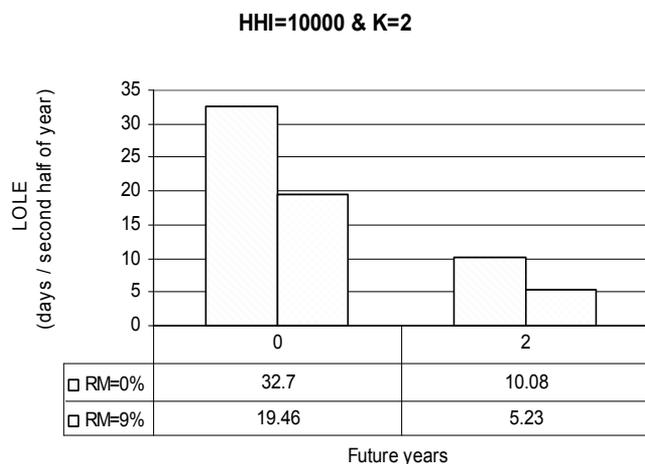


Fig. 19. *LOLE* values for the fourth study using N.N.

In all case studies, if reserve margin increases, *LOLE* will decrease and reliability will improve.

As mentioned before, in longer terms, customers can better adjust their load relative to the price. Therefore, price elasticity increases in longer terms, and according to (5), demand exponent curve reaches less gradient. As a result, intersection of the power plants' total offer curve and demand exponent curve will occur at less demand. This matter leads to operate from fewer power plants. Therefore, in each case study, if time increases, *LOLE* will decrease.

If market becomes more concentrated or *HHI* becomes bigger, *K* will find bigger value. Therefore, according to (5), intersection of the power plants' total offer curve and demand exponent curve will occur at less demand. Therefore, *LOLE* will decrease. So that in the fourth study (monopoly market), *LOLEs* are the least values comparing to the other case studies.

It is to be noted that since available capacity of hydro plants in IEEE-RTS are different in the first and the second halves of the year, therefore, simulations were done for the second half of the year. Evidently, the proposed method can be utilized for every simulation time. Also, in this study, it was supposed that the annual additional generation capacity is uniformly distributed between all the present generators.

## 5. Conclusion

This research deals with generation reliability assessment in power pool market using Monte Carlo simulation and intelligent systems. Since changes of market concentration in power markets are gradual, a fuzzy logic was proposed for calculation of the gradient coefficient of demand exponent curve. Due to the stochastic behavior of market and generators' FOR, MCS was used for the simulations. Also, for creation of a unique structure for reliability assessment, a neural network was used, which its outputs were very similar to MCS results. In this research, LOLE was used as reliability index and it was shown that if market becomes more concentrated, LOLE will decrease and reliability will improve. Also, if price elasticity of demand increases, LOLE will decrease.

Follows can be considered for future researches:

1. Reliability indices evaluate in HL-II zone in which both generation and transmission systems are considered.
2. Bilateral contracts consider in the power market as well as pool market.
3. If the generation planning scenarios in a power system are specified, then they can be used instead of uniformly distribution of annual additional generation capacity.
4. Reserve market can be considered as an independent market of the main energy market.

## 6. Symbol list:

MC: Marginal cost (mills/kWh)

MR: Marginal revenue (mills/kWh)

Q: Quantity of power (kW)

P: Electrical energy price (mills/kWh)

RM: Reserve margin (%)

$E_d$ : Price elasticity of demand (kW<sup>2</sup>h/mills)

$Q_n$ : Forecasted load (kW)

LOLE: Loss of load expectation (days/second half year)

FOR: Forced outage rate of power plants

$q_i$ : Share of  $i^{\text{th}}$  company in the pool market (%)

M: Number of independent companies in the market

$a$ : Demand exponent curve cross of basis (mills/kWh)

$b$ : Demand exponent curve gradient (mills /kW<sup>2</sup>h)

HHI: Hirschman - Herfindahl index

DE: Demand exponent curve

K: Gradient coefficient of demand exponent curve

MFU: Membership function of unconcentrated market

MFM: Membership function of moderately concentrated market

*MFH*: Membership function of highly concentrated market

*FT*: Simulated future time (year)

*NG*: Number of selected plants for generation in the market

*AGP*: Available generated power

## 7. References

- Billinton R., Allan R. (1992). *Reliability Evaluation of Engineering Systems*, Second edition, Plenum press, ISBN: 0-306-44063-6, New York.
- Billinton R., Allan R. (1996). *Reliability Evaluation of Power Systems*, Second edition, Plenum press, ISBN: 0-306-45259-6, New York.
- Borenstein Serverin (1999). Understanding competitive pricing and market power in wholesale electricity market, *University of California energy institute*.
- Draper E. L. (1998). Assessment of Deregulation and Competition, *IEEE Power Engineering Review*, Vol. 18, No. 7 (Jul 1998), pp. 17-18, ISSN: 0272-1724.
- Haroonabadi H. & Haghifam M.-R. (2009). Generation Reliability Evaluation in Power Markets Using Monte Carlo Simulation and Neural Networks. *Proceedings of 15<sup>th</sup> International Conference on Intelligent System Applications to Power Systems (ISAP)*, pp. 1-6, Print ISBN: 978-1-4244-5097-8, Curitiba, Nov 2009.
- International Energy Agency (IEA) (2002). *Security of Supply in Electricity Markets - Evidence and Policy Issues*, IEA, ISBN: 92-64-19805-9, France.
- International Energy Agency (IEA) (2003). *The Power to Choose- Demand Response in Liberalized Electricity Markets*, IEA, ISBN: 92-64-10503-4, France.
- Jaeseok Choi; Hongsik Kim; Junmin Cha & Roy Billinton (2001). Nodal probabilistic congestion and reliability evaluations of a transmission system under the deregulated electricity market, *Proceedings of IEEE Power engineering society summer meeting*, pp. 497-502, Print ISBN: 0-7803-7173-9, Vancouver, 15 Jul 2001-19 Jul 2001.
- Mc Clanahan R. H. (2002). Electric Deregulation, *IEEE Industry Application Magazine*, Vol. 8, No. 2 (Mar/Apr 2002), pp. 11-18, ISSN: 1077-2618 .
- Mountford J. D., Austria R. R. (1999). Keeping The Lights On, *IEEE Spectrum*, Vol. 36 (Jun 1999), pp. 34-39, ISSN: 0018-9235.
- Pindyck Robert S. & Rubinfeld D. L. (1995). *Microeconomics*, Third edition, Prentice Hall, ISBN: 7-302-02494-4, USA.
- Puttgen H. B.; Volzka D. R. & Olken M. I. (2001). Restructuring and Reregulation of The US Electric Utility Industry, *IEEE Power Engineering Review*, Vol. 21, No. 2 (Feb 2001), pp. 8-10, ISSN: 0272-1724.
- Reliability Test System Task Force of The IEEE Subcommittee on the application of probability Methods, IEEE Reliability Test System, *IEEE Transactions*, Pas-98, No.6, Nov/Dec 1979, pp. 2047-2054.
- Salvaderi L. (2000). Electric Sector Restructuring in Italy, *IEEE Power Engineering Review*, Vol. 20, No. 4 (Apr 2000), pp. 12-16, ISSN: 0272-1724.
- The U.S. Department of Justice and Federal Trade Commission (FTC) (1992). <http://www.ftc.gov/bc/docs/horizmer.htm>.

- Wang P. & Billinton R. (2001). Implementation of non-uniform reliability in a deregulated power market, *Proceedings of Canadian Conference on Electrical and Computer Engineerin.*, pp. 857- 861, Print ISBN: 0-7803-6715-4, Toronto, May 2001.
- Wang P.; Ding, Y. & Goel, L. (2009). Reliability assessment of restructured power systems using optimal load shedding technique, *Generation, Transmission & Distribution, IET*, Vol. 3, Issue: 7 (July 2009), pp. 628 - 640, ISSN: 1751-8687.

# Automating First- and Second-order Monte Carlo Simulations for Markov Models in TreeAge Pro

Benjamin P. Geisler, M.D., M.P.H.  
*United States of America*

## 1. Introduction

### 1.1 Note

This chapter draws heavily from a journal article (Geisler et al., 2009). The method described in there formed the basis for this book chapter and has been extended since.

### 1.2 Markov models

Markov models enable stochastic evaluation of linear and also non-linear decision problems that might otherwise be difficult to evaluate.

Named after the Russian mathematician Andrey Andreyevich Markov (1856-1922), Markov models enjoy a widespread appreciation in engineering, medicine, and other applied scientific disciplines.

For example, one application of this modelling approach lies in health technology assessment which aims to “forecast” clinical and economic consequences of adopting new diagnostic or treatment strategies. In this area, Markov models are typically used to estimate the long-term cost-effectiveness analysis of new medical interventions, i.e., weighing the incremental costs of a given intervention with its potential incremental health outcomes, and comparing their ratio to other incremental cost-effectiveness ratios (Weinstein & Stason, 1977).

However, Markov models are also useful to

- a. combine input parameters from different sources: short-term experimental studies (e.g., randomized controlled trials) with long-term observational studies, diagnostic with treatment strategies, costs with duration and quality-adjustment of life time, or any combination of these;
- b. extrapolate results to a longer time horizon, if possible ideally to lifetime;
- c. transfer or transpose data to a different patient cohort (e.g., a cohort with different baseline characteristics, a different compliance or adherence, or an important subgroup), to a different epidemiological situation (i.e., apply the model to other populations with other incidences and/or prevalences), to a different health care provider (who might follow a different standard of care), to a different payor (who might have different formula), or even to a different country or region.

Within a Markov model, often dubbed “state transition model”, the probability to be in one of the so-called Markov states always add up to one. The probability of changing states – or remaining in the same state, for that matter – depends in the classical Markov model solely

on the present state (Markovian property). This strict property can, however, be relaxed. Information *can* be maintained by “tracking” events as values in variables (tracking variables). Therefore, discrete events can be used to calculate state transition probabilities or rewards. The latter are used, for example, to calculate costs of health outcomes.

### 1.3 Monte Carlo simulations

As explained in more detail in other chapters of this book, there are two types of Monte Carlo simulations: first-order Monte Carlo simulations (sometimes referred to as “trials” or “microsimulations”) and second-order Monte Carlo simulations. These simulations both represent uncertainty that is not addressed in deterministic analysis (i.e., deterministic analysis does not use probabilities for parameters, just point estimates). However, first- and second-order Monte Carlo simulations represent different types of uncertainties that are consequently used for different purposes.

Second-order Monte Carlo simulation addresses parameter-uncertainty. A typical application in health care modelling is probabilistic sensitivity analysis of cost-effectiveness models. Probabilistic sensitivity analysis assesses the “joint effect of parameter uncertainty” (Briggs et al., 2002) on the result of the model, i.e. when randomly drawing from all distributions simultaneously. In cost-effectiveness analysis, the thus computed incremental costs and health outcomes are most commonly plotted on an X-Y scatter plot. Another very common depiction of the results of a probabilistic sensitivity analysis are cost-effectiveness acceptability curves where the “willingness to pay” (e.g., in dollars per quality-adjusted life year) is plotted against the proportion of runs that resulted in incremental cost-effectiveness ratios below this willingness to pay (Fenwick et al., 2004).

First-order Monte Carlo simulation addresses stochastic uncertainty. Their application is less straightforward.

However, in a more relevant application, first-order Monte Carlo simulations employing “tracker variables” enable one to overcome the Markov property. Another way to do this would be to save information by increasing the number of health states. However, “state explosion” makes model-building more prone to errors. Tracker variables offer an attractive way to incorporate discrete events into Markov models.

Moreover, first-order Monte Carlo simulations might be an interesting way to reflect heterogeneity in patient populations, a desirable goal.

Stochastic uncertainty from a practical point of view is considered “random noise from a decision maker’s point of view” but, as a means to an end, “can be overcome by increasing the sample size of the microsimulation” (Weinstein, 2006) to achieve the above stated goals.

The use of first-order Monte Carlo simulations in decision-analytic modelling has constantly increased over time (Hunink et al., 2001). This is largely due to the fact that drastically increased computer power has made it possible to perform microsimulations in reasonable amounts of time.

However, with a standard modelling package such as TreeAge Pro™, it can be difficult to perform multiple deterministic sensitivity analyses in model requiring first-order Monte Carlo simulations, since each single data point requires another microsimulation to be performed, and thus huge investments of time have to be spared by the researcher. The goal of this chapter is therefore to identify ways to automate deterministic one-way sensitivity analysis of models requiring the use of first-order Monte Carlo simulations.

## 1.4 TreeAge Pro™

Formerly named DATA Pro, TreeAge Pro™ (TreeAge Software, Inc., Waltham, MA, USA) is popular modelling software capable of computing expected values of Markov models as well as deterministic sensitivity, threshold and probabilistic sensitivity analysis via second-order Monte Carlo simulation. TreeAge Pro™ is also able to run microsimulations via first-order Monte Carlo simulation. Advanced value-of-information analysis features in TreeAge include expected value of partially perfect information analysis via two stacked second-order Monte Carlo simulations.

While value-of-information analysis based on microsimulation models can be run “out of the box” in three-dimensional models, it is not possible to run simple deterministic sensitivity analysis on models requiring microsimulations in TreeAge Pro™.

A relatively new analysis type, cost-effectiveness sensitivity curves, requires repeated runs of second-order Monte Carlo simulations (O’Day et al., 2010). This is also not natively supported by TreeAge Pro™.

## 1.5 Chapter objectives

This chapter will explain the techniques to automatically run Monte Carlo simulations, in TreeAge Pro, which potentially saves decision scientists valuable time and energy. The text will provide the open-source codes for Excel®- and Visual Basic®-based interfaces to run these analyses as well as to adapt and develop them further.

## 2. General materials and methods

### 2.1 Object-oriented programming

In general, the goals of Object-Oriented Programming are (1) to re-use programs and sub-programs; and (2) to ease the use by allowing programs to be accessed by other programs. In Object-Oriented Programming, *objects* are defined classes, methods, or other “building blocks”. That means that objects can be initiated and reused by other objects (created, accessed by other programs). Objects have certain behaviours (the things it can do, or features), and objects will have certain characteristics (attributes, fields, parameters, data, or properties). They can exist in a certain hierarchy where a daughter instance “inherits” all objects from the mother instance. This again eases the use of the coder as he/she does only need to define objects once and for all.

### 2.2 TreeAgeProLib

TreeAge Pro™ features graphical user interface, TreeAge Pro™ as we know it. However, starting with version 2007, TreeAge Pro™ also includes an object or “scripting” interface which provides access from other programs.

TreeAge Pro Suite/Excel installs an “add-in” to Excel that adds TreeAge Pro™ objects to be accessed by Excel (see below for preparations in Excel® and Visual Basic®).

“TreeAgeProLib” is an object library that implements a standard COM interface to create and access TreeAge Pro™ objects in any program, script, or macro can create or access TreeAge Pro™ objects through this interface.

An object interface enables automating of tasks that are common, repetitive, and time-consuming, e.g., (re)setting a tree’s variables from outside TreeAge Pro™, automatically running a set of analyses, or automatically exporting results. The object interfaces integrate

TreeAge Pro™ functions into other systems or applications like Excel® via Visual Basic®, into websites (via Microsoft®'s .ASP technology), or into models programmed for examples in C++.

The object browser is included in the Microsoft Visual Basic® for Excel® editor and shows object properties and methods. The objects available can be browsed in Visual Basic® by choosing Object Browser from the “View” menu or by simply pressing <F2>. After choosing the TreeAgeProLib, one can browse as well as search all objects and see descriptions and syntaxes (please see Figure 1).

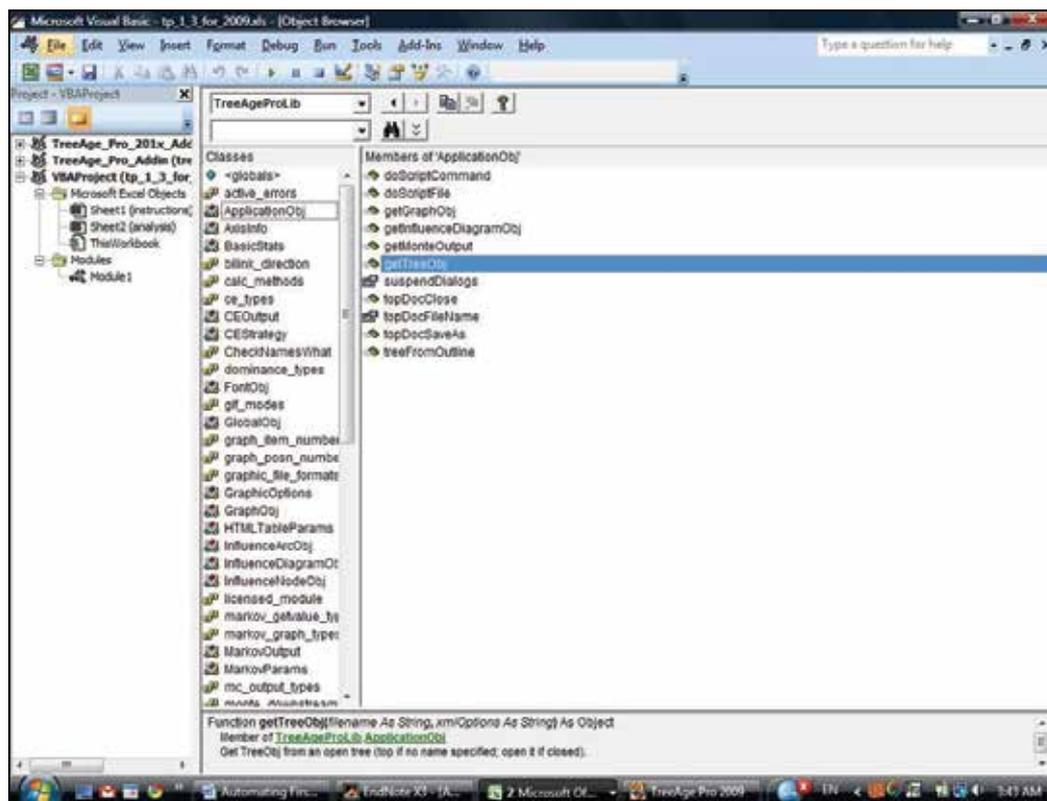


Fig. 1. Browsing objects from TreeAgeProLib object library in Visual Basic®

### 2.3 Software requirements

The object interface requires the Excel® Module (included in Tree Age Pro™ Excel and Tree Age Pro™ Suite). Tree Age Pro™ Suite 2009 includes both the Healthcare and the Excel® packages which are necessary for our purposes. All analyses have been conducted in TreeAge Pro™ Suite 2009 (from now on simply referred to as TreeAge Pro). The model will also open and the Visual Basic®/Excel® interface will function in earlier versions. This has been successfully tested in TreeAge Pro™ 2007 and 2008. However, the interface will not work with adaptations in the soon-to-be-released TreeAge Pro™ 2011. The reason for this is that TreeAge Pro™ 2011 uses a completely different platform, ECLIPSE™ integrated development environment, based on the programming language Java.

## 2.4 Preparations in Excel® and Visual Basic®

Please note that Excel® needs to be able to execute macros. You can tell Excel® whether it shall accept macros at Tools > Options > Security > Macro security. In order to execute our macro, you need to put the macro security level to medium (or low). You will then be prompted about whether you want to allow macros when opening the spreadsheet file.

The Visual Basic® editor can be opened in Excel® by pressing Alt + F11. In Microsoft Office® Excel® versions 2007 and 2010, the editor can be reached via the developer tab in the “ribbon”.

First, the TreeAge Pro Suite add-on in Excel® needs to be enabled. This adds TreeAge objects that can then be accessed by Excel®. Click on Tools > References and make sure that the reference “TreeAge Pro 2009.0.0 Type Library” is enabled (please see Figure 2).

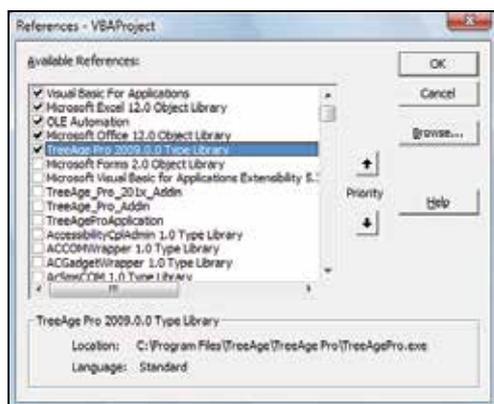


Fig. 2. References in the Visual Basic® Editor

In the tree view on the left, click then on “Module1” (under VBAProject (filename) > Modules) if it is not already opened in the editor.

## 3. Automating deterministic one-way sensitivity analysis in Markov Models requiring first-order Monte Carlo simulations

Next we tackle the generic script that was constructed which enables the performance of automated deterministic one-way sensitivity analyses in Excel employing microsimulation models.

We use object-oriented programming to access the TreeAgeProLib objects of the TreeAge Pro Excel® add-in to perform a specified number of first-order Monte Carlo simulations in Excel®.

The code in Visual Basic® is given below in the next subsection 3.1. In the subsequent section 4 we will then discuss an easy way to access the script.

### 3.1 Visual Basic® code

The goal of the script is to run a file in TreeAge repeatedly while varying one parameter during each run, and subsequently output the results.

More specifically, the coding objectives are to

1. open the tree;
2. loop for the purposes of running model for each value that the sensitivity analysis requires;

3. repeatedly manipulate variable in root;
4. repeatedly run the Monte Carlo simulation; and
5. output effectiveness, costs, incremental cost-effectiveness ratio in Excel®.

We will use the following TreeAgeProLib object:

- *ApplicationObj*, used to open the tree
- *TreeObj*, to access the tree
- *NodeObj*, to navigate to the route and change the parameter of interest
- *MonteParams*, to set up the Monte Carlo simulation
- *MonteOutput*, to access the results of the Monte Carlo simulation
- *BasicStats*, to access the costs and effectiveness results

Given below is a sample code that needs to be put into the Visual Basic® for Excel® editor.

Please note that every line that starts with an apostrophe (') and in the colour green is just a comment line and not necessary for the program.

Please also note that terms in italics need to be replaced with either the name of the appropriate worksheet or the appropriate cell number for the input parameters.

```

Sub one_way_w_two_strategies()
' the locations and name of our model; "" means that the currently opened tree is being used
Const pkgname = ""
' number of microsims
Dim runs As Double
' start row and current row
Dim srow As Integer
Dim currow As Double
' variable of interest
Dim var As String
' startvalue, number and value of interval (between the values)
Dim nInter As Integer
Dim vInter As Double
Dim sval As Double
' seeding
Dim seedN As Long
Dim seedB As Integer
' declaration of variables for use within the loop
Dim i As Integer
Dim cur As Double
Dim cur_string As String
Dim root As NodeObj
Dim mcParams As New MonteParams
Dim mcOutput As MonteOutput
Dim cstats_nt As BasicStats
Dim cstats_t As BasicStats
Dim estats_nt As BasicStats
Dim estats_t As BasicStats
Dim err2 As Long

```

```

Dim icer As Double
Dim c_nt As Double
Dim c_t As Double
Dim e_nt As Double
Dim e_t As Double
Dim c_i As Double
Dim e_i As Double
Dim curcolumn As Integer
Const save = 0
'first a message
    Application.ScreenUpdating = False
    ' turns off screen updating
    Application.DisplayStatusBar = True
    ' makes sure that the statusbar is visible
    Application.StatusBar = "Please wait while the tree is being opened..."
'definition of number of microsimulations
ActiveWorkbook.Sheets("name of worksheet ").Select
Range("appropriate cell for number of runs").Select
runs = ActiveCell.Value
'definition of start row
srow = 20
currow = srow
'definition of variable of interest
ActiveWorkbook.Sheets("name of worksheet ").Select
Range("name of worksheet for variable of interest").Select
var = ActiveCell.Value
'seeding
Range("name of worksheet for seed number").Select
seedN = ActiveCell.Value
Range("name of worksheet for seed behavior").Select
seedB = ActiveCell.Value
'the following code lines serve the purpose to read the start value, number of intervals and
value of intervals from the spreadsheet; if not the spreadsheet from the website is being
used, the cell values might have to be adjusted
Range("name of worksheet for starting value").Select
sval = ActiveCell.Value
Range("name of worksheet for number of intervals").Select
nInter = ActiveCell.Value
Range("name of worksheet for the value of each interval").Select
vInter = ActiveCell.Value
'start of the loop
For i = 1 To nInter
    'definition of the current value of the variable of interest; determined by the loop
    number (i)
    cur = sval + ((i - 1) * vInter)
    cur_string = Str$(cur)

```

```

'initiation of application and tree
Dim appObj As TreeAgeProLib.ApplicationObj
Dim tree As TreeAgeProLib.TreeObj
Set appObj = New TreeAgeProLib.ApplicationObj
Set tree = appObj.getTreeObj(pkgname, "")
Application.Wait Now + TimeValue("00:00:01")
Application.StatusBar = "The tree is opened. Now the MC simulation will be run..."
'manipulation the root and setting the variable of interest to the current value
Dim ok As Long
ok = tree.setVariableValue(defined_at_root, var, cur_string)
'monte carlo microsimulation/ trials (1st order monte carlo simulation) and threads
mcParams.trials = runs
mcParams.sampleType = 0
mcParams.seedValue = seedN
mcParams.seedBehavior = seedB
Application.StatusBar = False
Set mcOutput = tree.MonteCarlo(mcParams)
While mcOutput.TimeElapsed = 0
    Application.Wait Now + TimeValue("00:00:02")
Wend
'check
If mcOutput.valid = 0 Then
    MsgBox ("McOutput invalid")
    Exit Sub
End If
' costs and efficiency by strategy
Set cstats_nt = mcOutput.GetStatsByStrategy(ce_cost, 1, 0.1)
Set estats_nt = mcOutput.GetStatsByStrategy(ce_eff, 1, 0.1)
Set cstats_t = mcOutput.GetStatsByStrategy(ce_cost, 2, 0.1)
Set estats_t = mcOutput.GetStatsByStrategy(ce_eff, 2, 0.1)
'calculate icer and output
c_nt = cstats_nt.mean
c_t = cstats_t.mean
e_nt = estats_nt.mean
e_t = estats_t.mean
c_i = c_t - c_nt
e_i = e_t - e_nt
If e_i <> 0 Then icer = c_i / e_i Else icer = 0
If icer < 0 Then icer = 0
'definition of the current row
currow = srow + (i - 1)
'output
curcolumn = 1
ActiveWorkbook.Sheets("name of worksheet ").Cells(currow, curcolumn).Value = cur
curcolumn = curcolumn + 1
ActiveWorkbook.Sheets("name of worksheet ").Cells(currow, curcolumn).Value = icer

```

```

curcolumn = curcolumn + 1
ActiveWorkbook.Sheets("name of worksheet ").Cells(currow, curcolumn).Value = c_i
curcolumn = curcolumn + 1
ActiveWorkbook.Sheets("name of worksheet ").Cells(currow, curcolumn).Value = e_i
curcolumn = curcolumn + 1
ActiveWorkbook.Sheets("name of worksheet ").Cells(currow, curcolumn).Value = c_nt
curcolumn = curcolumn + 1
ActiveWorkbook.Sheets("name of worksheet ").Cells(currow, curcolumn).Value = c_t
curcolumn = curcolumn + 1
ActiveWorkbook.Sheets("name of worksheet ").Cells(currow, curcolumn).Value = e_nt
curcolumn = curcolumn + 1
ActiveWorkbook.Sheets("name of worksheet").Cells(currow, curcolumn).Value = e_t
  'reset of all variables
  Set mcOutput = Nothing
  Set tkt = Nothing
  Set tree = Nothing
  Set root = Nothing
  Set mcParams = Nothing
  Set mcOutput = Nothing
  Set cstats_nt = Nothing
  Set cstats_t = Nothing
  Set estats_nt = Nothing
  Set estats_t = Nothing
  err2 = 0
  'closure of the Monte Carlo simulation window
  appObj.topDocClose (0)
Next
End Sub

```

---

### 3.2 Excel® spreadsheet

In addition to the script, a spreadsheet was constructed in Excel® (please see Figure 3). The workbook allows one to make use of the script without programming knowledge.

The script and spreadsheet (as well as the script that comes with it) is generic for all TreeAge decision trees with two strategies and Markov nodes

The spreadsheet, which includes the script as a macro, can be downloaded from the author's website [www.hcval.com](http://www.hcval.com).

## 4. Automating cost-effectiveness sensitivity curves in Markov Models requiring second-order Monte Carlo simulations

Analogous to the previous section, let us explore if there are applications that make it worthwhile to also automate second-order Monte Carlo simulations and how they can be realized.

One possible application is expected value of partially perfect information analysis, which requires either two stacks of second-order Monte Carlo simulations, or even, if first-order Monte Carlo simulation is required to derived expected values, three levels of analysis.

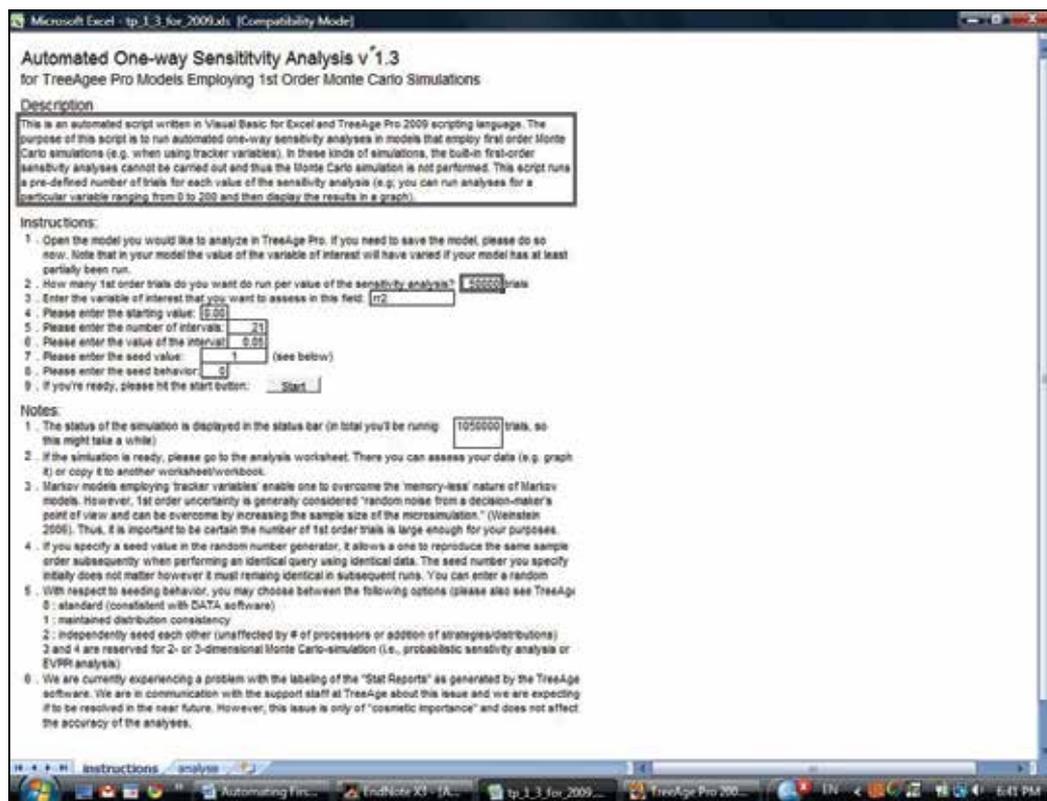


Fig. 3. Generic Excel® spreadsheet

However, both are natively supported in TreeAge Pro and these features generally perform well.

Another possible application are cost-effectiveness sensitivity curves (O'Day et al., 2010). Cost-effectiveness sensitivity curves are a combination of one-way sensitivity analysis and probabilistic sensitivity analysis. For each curve, one willingness-to-pay level needs to be specified, let us say by convention \$50,000/QALY and \$100,000/QALY. One parameter of interest is varied and, for example, plotted on the x-axis. For each value that this parameter can take in the range specified on the x-axis, a separate second-order Monte Carlo simulation has to be run. In this type of analysis, we plot the proportion or percentage of runs that were at or below the specified willingness to pay threshold on the y-axis.

It would be desirable to run these kinds of analyses either in normal deterministic Markov models or in those requiring First-order Monte Carlo simulations (because tracker variables are used). Unfortunately, the methods to extract the necessary values from second-order Monte Carlo simulations are currently not specified for the object in question, MonteOutput. Hopefully, in the future this will be supported by TreeAge Pro.

## 5. Example

Consider a simple example: a Markov model with two states, alive and dead, that nevertheless represents a chronic diseases that has a higher probability of recurrence and a worse quality of life (morbidity) as well as a higher risk of death (mortality) with each

recurrence. Additionally, the treatment gets more and more expensive as the disease progresses.

Please see Figures 4 through 6 for representations of the model.

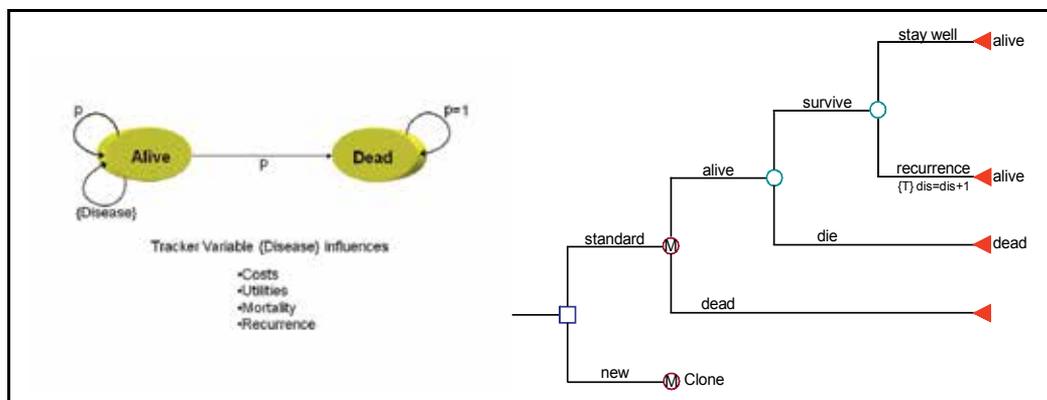


Fig. 4. Illustrative Markov Model requiring a tracker variable (bubble diagram and decision tree with Markov states)

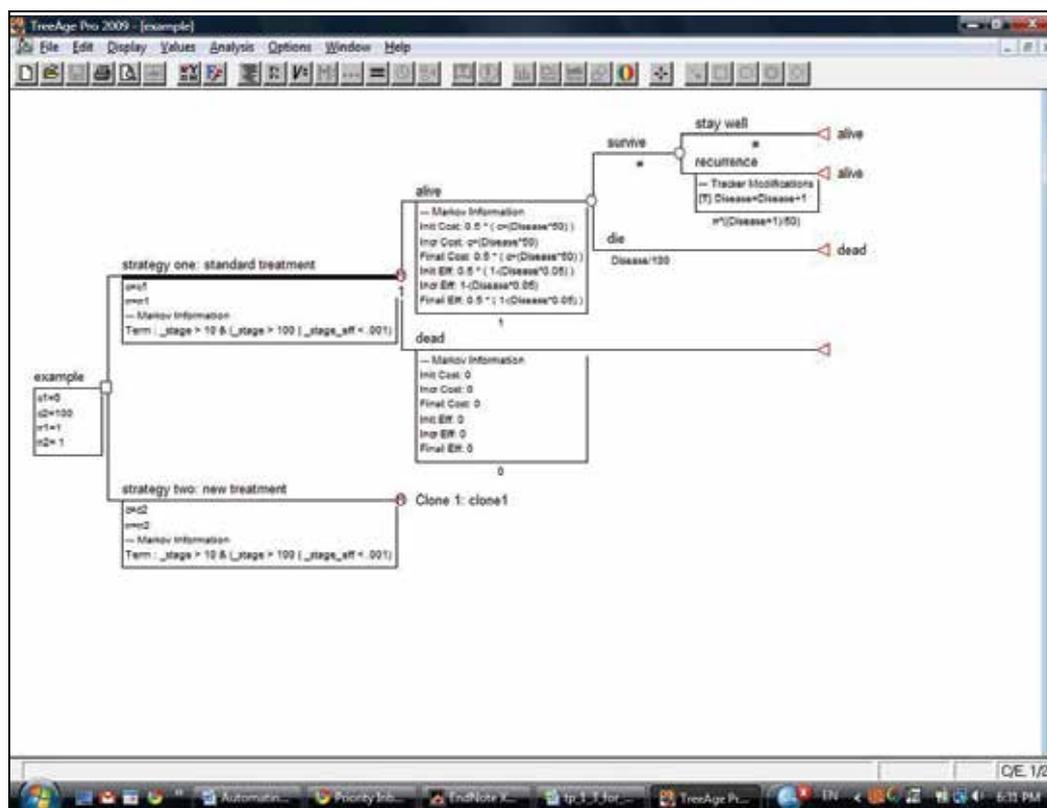


Fig. 5. Illustrative Markov Model requiring a tracker variable (screenshot)

Although this is such a minimalistic model, it is possible to add complexity by other means: e.g., a third state, recurrence, could be modelled within the alive state when a tracker variable, let us call it {Disease}, assumes a certain value. Every time a patient gets a recurrence, the value of the tracker variable goes up by one. The tracker variable then in turn influences the transition probability of the recurrence, the transition probability to death (mortality), costs (severer disease might be more expensive), and utility (multiple recurrences might be associated with lower health-related quality of life). Please see Table 1 for the actual variable definitions in this example.

Parameter	Strategy 1 - standard -	Strategy 2 - new -
Costs	Disease*50	100+Disease*50
Utilities	1-(Disease*0.05)	
Recurrence	$\frac{\text{Disease}+1}{25}$	$\frac{RR*(\text{Disease}+1)}{25}$
Mortality	$\frac{\text{Disease}}{100}$	

Table 1. Input parameters for the example

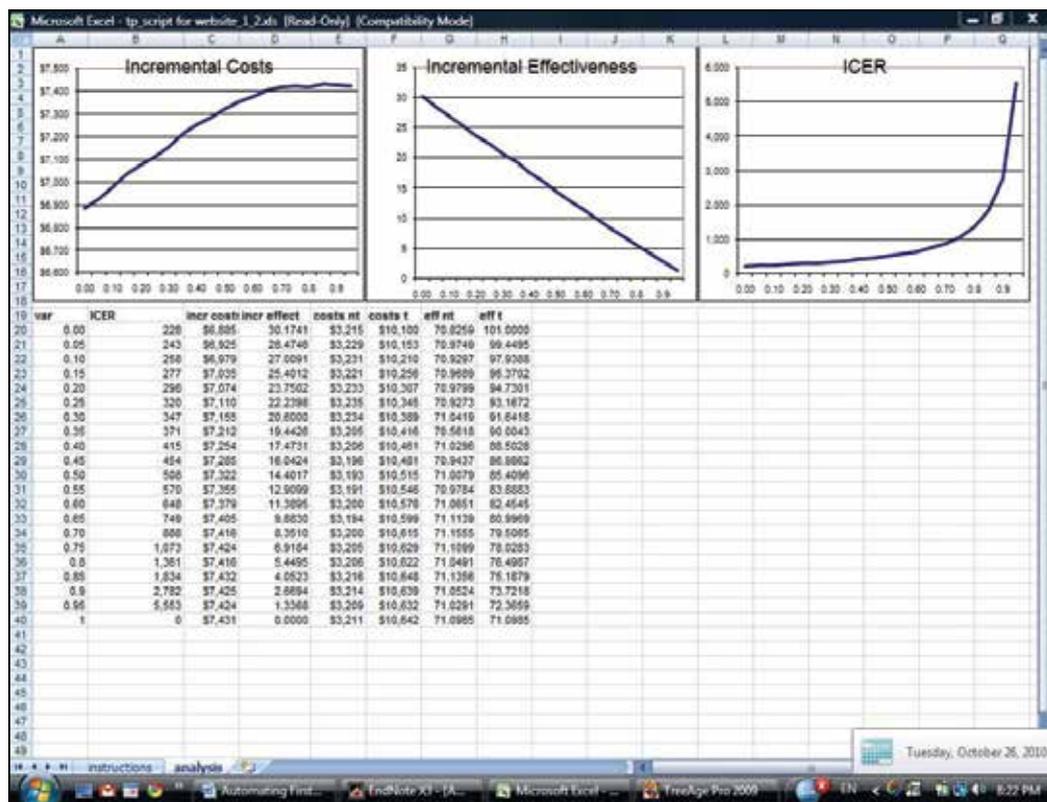


Fig. 6. Resulting deterministic sensitivity analysis from the example in Excel®

In this example, we chose to perform a deterministic one-way sensitivity analysis on  $rr_2$ , i.e., we vary the relative risk that describes how a new treatment (strategy 2) influences the risk of recurrence. Since we are using a tracker variable, we need to perform first-order Monte Carlo simulations *for the entire range of values* that we chose to analyze.

A screenshot of the results in Excel® is provided in Figure 6. As can be seen, the lines are not smooth or (in case of a linear function) straight. This is the result of the model relying on first-order Monte Carlo simulations. However, with increasing sample size, the smoothness of the lines should approach the level of deterministic analysis.

## 6. Other applications

A relatively new analysis type, cost-effectiveness sensitivity curves, requires repeated runs of second-order Monte Carlo simulations (O'Day et al., 2010). This is neither natively supported by TreeAge Pro, nor – as explained above – can this analysis type be run via an automated script from Excel®/Visual Basic®.

Other, potentially interesting applications are automated sensitivity analysis of cohort models, automated deterministic two- or multi-way-sensitivity analyses of models depending on first-order Monte Carlo simulations, automated model calibration, and automated second-order Monte Carlo simulations such as expected value of partial information analyses with varying sets of constant variables.

The online documentation is available at:

<http://server.treeage.com/ObjDocs/TP/TOC/ref.php3> and includes descriptions of all objects and their properties and methods as well as code samples.

Please check the website [www.hcval.com](http://www.hcval.com) for updates.

## 7. Conclusions

Linking TreeAge Pro and Microsoft Excel® via Microsoft Visual Basic® allows the automatic performance of multiple deterministic sensitivity analysis in first-order-Monte Carlo simulations.

The automatic deterministic sensitivity analysis in models requiring first-order-Monte Carlo simulations can be run conveniently via Excel® spreadsheets, with no programming knowledge necessary. A sample worksheet in Excel® for all TreeAge decision trees with two strategies and Markov nodes allows for use of the script without programming knowledge and can be downloaded from the website [www.hcval.com](http://www.hcval.com).

However, it is currently impossible to automate second-order Monte Carlo simulations in any meaningful ways. An example where this would be helpful is the automatic calculation of cost-effectiveness sensitivity curves based on multiple probabilistic sensitivity analyses based on second-order Monte Carlo simulations (O'Day et al., 2010).

## 8. Acknowledgements

The origin of this work was in a decision-analytic model that the author created together with Alexander Göhler, MD PhD MSc MPH, Uwe Siebert, MD MPH MSc ScD, G. Scott Gazelle, MD MPH PhD, and David J. Cohen, MD MSc. The author is indebted to these individuals for accepting him as a research fellow to perform this research, and for co-authoring a journal article on the method that this decision model used and that formed the basis for this book chapter (Geisler et al., 2009). The book chapter draws heavily on the work accomplished in those previous projects.

The author would also like to thank TreeAge Software, Inc. Support for help with new features which were not yet documented. The author is not affiliated with TreeAge Software, Inc.

Software and trademarks are the property of the respective copyright holders. In particular, "TreeAge Pro" and "DATA" and the TreeAge Software logo are TreeAge Software, Inc. trademarks. ECLIPSE is a trademark of Eclipse Foundation, Inc. Java is a registered trademark of Oracle and/or its affiliates. Microsoft® is a registered trademark of Microsoft Corporation in the United States and other countries. Microsoft Office® is a registered trademark of Microsoft Corporation in the United States and other countries. Microsoft Excel® is a registered trademark of Microsoft Corporation in the United States and other countries. These trademarks and trade names are the property of their respective owners. Other trademarks or trade names may be the property of their respective owners.

## 9. References

- Briggs, A.H., Goeree, R., Blackhouse, G. & O'Brien, B.J. (2002). Probabilistic analysis of cost-effectiveness models: choosing between treatment strategies for gastroesophageal reflux disease. *Med Decis Making*, 22, 4, (Jul-Aug 2002) 290-308, ISSN 0272-989X.
- Fenwick, E., O'Brien, B.J. & Briggs, A. (2004). Cost-effectiveness acceptability curves--facts, fallacies and frequently asked questions. *Health Econ*, 13, 5, (May 2004) 405-415, ISSN 1057-9230.
- Geisler, B.P., Siebert, U., Gazelle, G.S., Cohen, D.J. & Gohler, A. (2009). Deterministic sensitivity analysis for first-order Monte Carlo simulations: a technical note. *Value Health*, 12, 1, (Jan 2009) 96-97, ISSN 1098-3015.
- Hunink, M.G.M., Glasziou, P.P., Siegel, J.E., Weeks, J.C., Pliskin, J.S., Elstein, A.S. & Weinstein, M.C. (2001). *Decision making in health and medicine : integrating evidence and values*. Cambridge ; New York, Cambridge University Press.
- O'Day, K., Meissner, B. & Bramley, T. (2010). The Cost-effectiveness Sensitivity Curve: Quantifying the Effect of Individual Parameter Uncertainty in a Probabilistic Model. ISPOR Fifteenth Annual International Meeting. International Society for Pharmacoeconomics and Outcomes Research. Atlanta, GA, Value in Health. 13; 3: A1-219.
- Weinstein, M.C. (2006). Recent developments in decision-analytic modelling for economic evaluation. *Pharmacoeconomics*, 24, 11, (2006) 1043-1053, ISSN 1170-7690.
- Weinstein, M.C. & Stason, W.B. (1977). Foundations of cost-effectiveness analysis for health and medical practices. *N Engl J Med*, 296, 13, (Mar 31 1977) 716-721, ISSN 0028-4793.

# Monte Carlo Simulations of Adsorbed Molecules on Ionic Surfaces

Abdulwahab Khalil Sallabi  
*Misurata University*  
*Libya*

## 1. Introduction

Monte Carlo (MC) method [1-10] refers to all calculations that involve the use of random numbers for sampling processes of approximate solutions to quantitative problems. It can be applied for application domains range from economics to physics to chemistry to surface science to medicine.

The Monte Carlo (MC) method is usually Linked to Comte de Buffon a French eighteenth-century naturalist, who performed an experiment by throwing a needle of length  $\ell$  at random onto a board marked with parallel lines a distance  $d$  apart to infer the probability  $p$  that the needle will intersect one of those lines. Buffon's subsequent experiments enabled him to make an accurate estimation of  $\pi$ . Following the procedure of Buffon, Laplace, and then In 1864, Captain O. C. Fox and in 1873, A. Hall [8] used Monte Carlo method calculate  $\pi$ .

Early 1940's marked the beginning of the modern history of Monte Carlo when scientists at Los Alamos systematically used them as a research tool in their work on developing nuclear weapons. Stanislaw Ulam was the first one to realize the potential of using computers to automate the statistical sampling process. Stanislaw Ulam, John von Neuman and Nicolas Metropolis developed algorithms and explored the means to convert non-random problems into random forms so that statistical sampling can be used for their solution. The name "Monte Carlo" was suggested by Metropolis after the famous Monaco casino. One of the first published papers on this topic was by Metropolis and Ulam in 1949 [9].

## 2. Metropolis Monte Carlo method

Nicolas Metropolis introduced the Metropolis Monte Carlo method at the dawn of the computer era in 1953 [10]. The rapid development in computer technology has increased the applicability and accuracy of the Monte Carlo method and is now used routinely in many diverse fields, such as economics, physics and chemistry as a powerful numerical technique. As we know, it is easy to solve equations of interaction between two atoms or molecules and get an exact solution for a specific problem while in the case of large systems, where the number of particles involved in a problem is large, solving the problem in a deterministic way becomes impossible due to the large number of equations and variables that are needed to study the problem. Monte Carlo (MC) methods are stochastic (random) techniques in which random numbers and probability statistics are used to examine scientific problems in a probabilistic fashion rather than a deterministic one. To study the physical properties of a

system with a large number of atoms or molecules interacting with each other, MC methods can be readily applied whereby possible configurations of the system can be sampled according to their Boltzmann probability distribution via the use of random numbers [1] [3] [11] [12].

Metropolis Monte Carlo simulations have been performed to study the structures and phase transitions of adsorbed molecules on solid surfaces such as HBr/LiF(001) [13], CO<sub>2</sub>/NaCl [14], CO/NaCl [15], CO/LiF [16], CO/MgO [17], N<sub>2</sub>/NaCl [18], N<sub>2</sub>/LiF [19], H<sub>2</sub>/NaCl [20], D<sub>2</sub>/MgO [21], H<sub>2</sub>/LiF [22,23]. They have also been used to study critical phenomena near their transition temperatures for many models such as the Ising, XY, and Heisenberg models [12]. MC methods have proved to be useful tools since they allow for the sampling of a large number of possible configurations at nonzero temperatures.

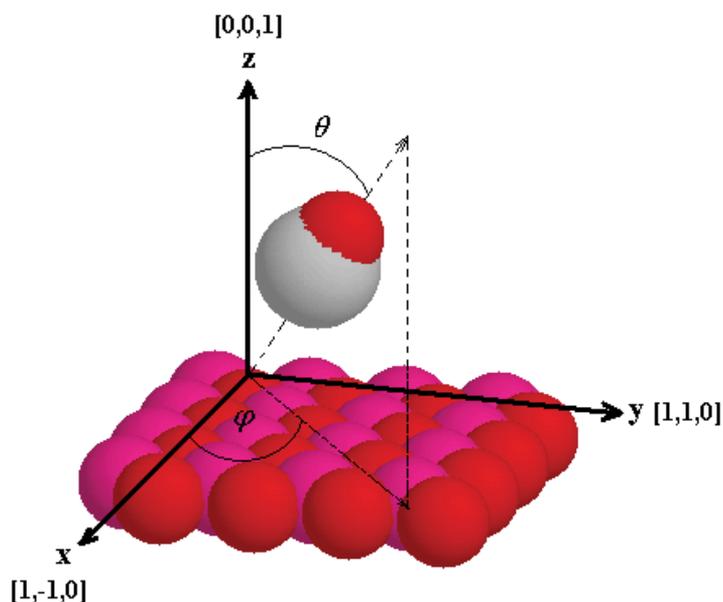


Fig. 1. A view of the angular coordinate system with respect to the xyz coordinate system. The polar angle  $\theta$  is the tilt of the molecular axis (carbon to oxygen) with respect to the surface normal (z-axis) while the azimuthal angle  $\varphi$  is the angle between the x-axis and the projection of the molecular axis onto the plane of the surface (xy-plane).

By using the Metropolis Monte Carlo (canonical ensemble or  $Q(N,V,T)$ ) technique [10], we can simulate the interaction of adsorbed molecules on ionic surfaces, where the number of molecules  $N$ , the simulation volume  $V$ , and temperature  $T$  are fixed during any simulation. In other words, during a simulation the positions and the orientations of molecules are allowed to change while the number of molecules, volume and temperature are not allowed to change. A MC simulation is typically broken down into cycles. In every cycle, each adsorbed molecule is allowed to move in a random fashion. In each move a randomly chosen molecule is moved to a new random position or orientation. Then the computer decides whether to accept or reject this move with a Boltzmann probability  $\exp(-\Delta E / k_B T)$  that depends on the change in energy ( $\Delta E = E_{\text{new}} - E_{\text{old}}$ ) of the new ( $E_{\text{new}}$ ) and old ( $E_{\text{old}}$ ) configuration. This process is repeated many times until there is no further change in the

average energy and other computed properties of the system, at which point the system is deemed to have reached thermodynamic equilibrium. After this point is reached, a large number of configurations (geometries) are accumulated and the data are averaged to obtain thermodynamic properties of the system, such as the energy and angular distributions. An example of one of the coordinate systems which can be used to describe the adsorbed molecules is shown in Fig. 1, in this coordinate system the position of an adsorbed molecule is described by the position vector  $\mathbf{r}(x,y,z)$  with respect to the origin which is taken in the plane  $z=0$  (the surface of the substrate) and at the anion site with the  $x$  and  $y$  axes running along the  $[1,-1,0]$  and  $[1,1,0]$  crystallographic directions respectively and with the  $z$  axis set perpendicular to the surface. The orientation of an adsorbed molecule is described by a polar angle  $\theta$  and an azimuthal angle  $\varphi$ . In Fig. 2 we show the general steps of the Metropolis Monte Carlo simulation [24].

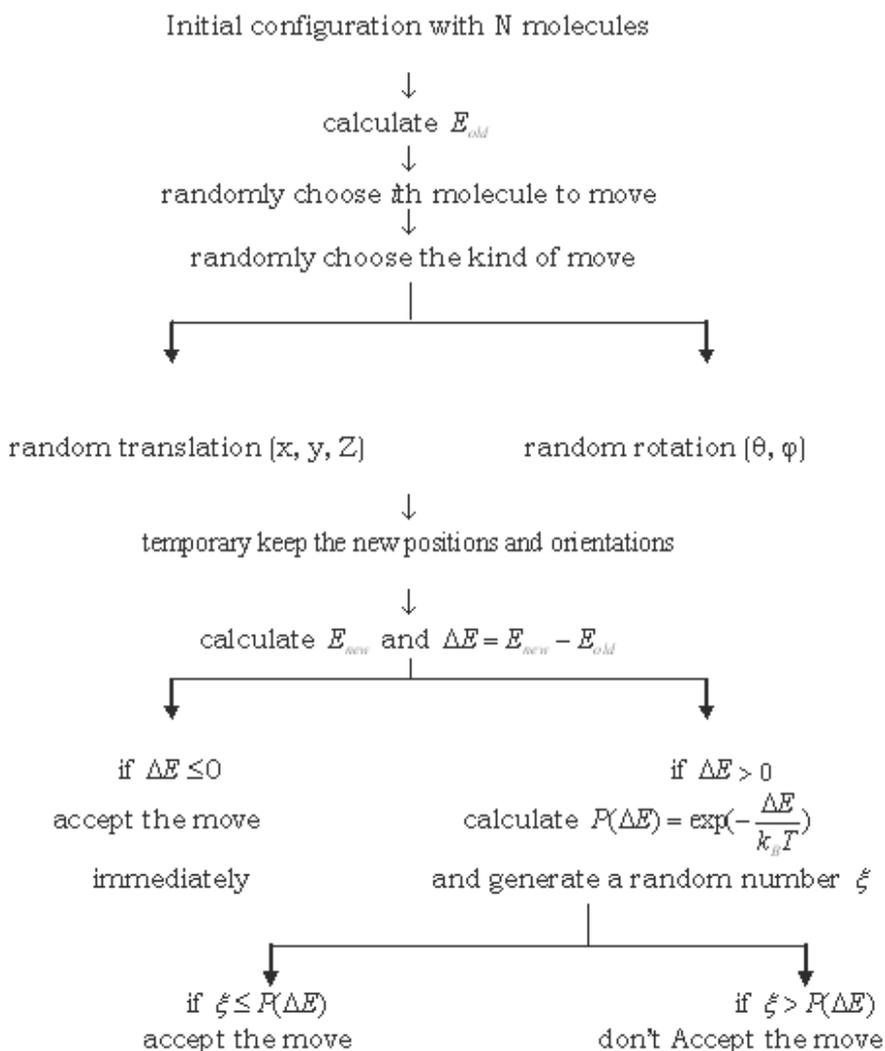


Fig. 2. Flow chart of general steps of Metropolis Monte Carlo method

During the simulations an ensemble of  $N$  adsorbed molecules are placed in the surface potential of the ionic substrate. The surface potential is fixed in the space. If all of the positive sites of the surface are occupied then a monolayer of adsorbed molecules is formed. Periodic boundary conditions in the lateral directions ( $x$  and  $y$ ) were imposed as well as a cutoff radius for the molecule-molecule interactions. Each move is subjected to the usual Boltzmann weighted acceptance criterion. The amplitudes of the moves were adjusted independently and maintained a 50% acceptance probability. Statistics were collected after the system is equilibrated. In the case of studying critical phenomena, analysis near the critical temperature needs very long runs to obtain good results.

### 3. Statistics

In order to study the continuous phase transition expected for the some of the adsorbed molecules on ionic surface (e.g.,  $N_2$  on  $NaCl$ .) we need to calculate statistically the average energy ( $E$ ), heat capacity ( $C_v$ ), order parameter ( $\Phi$ ) and susceptibility ( $\chi$ ) for a monolayer of adsorbed molecules on a square lattice with  $L \times L = N$  sites. To do that the order parameter and the energy per molecule were collected every cycle and have been kept for further analysis. The heat capacity per particle is obtained from the fluctuations in the monolayer energy  $E$  via [24],

$$C_v = (\langle E^2 \rangle - \langle E \rangle^2) / Nk_B T^2 \quad (1)$$

The order parameter  $\Phi$  tells us how well the system is ordered. When the system is perfectly ordered the order parameter has a value of one while it has a value of zero when the system is fully disordered. The order parameter  $\Phi$  for an anti-ferromagnetic like ground state is defined in terms of the azimuthal angle  $\varphi$  through the following relations and transformations [25].

$$\Phi = \sqrt{\Phi_x^2 + \Phi_y^2} \quad (2)$$

$$\Phi_x = N^{-1} \sum_{i=1}^N (-1)^{n_{yi}} \cos(\varphi_i) \quad (3)$$

$$\Phi_y = N^{-1} \sum_{i=1}^N (-1)^{n_{xi}} \cos(\varphi_i) \quad (4)$$

where  $n_{xi}=1,2,3,\dots,L$  and  $n_{yi}=1,2,3,\dots,L$ , label the  $x$  and  $y$  positions of the adsorption sites of molecule  $i$  on the lattice. The summations are taken over all the localized adsorption sites of the molecules on a square lattice from 1 to  $N=L^2$ . The transformed order parameter [146]  $\Phi$  is obtained by using the phase factors  $(-1)^{n_{xi}}$  and  $(-1)^{n_{yi}}$ . If the configuration of the adsorbed molecules is of the antiferro-type then the transformed configuration will be ferro-type where all the adsorbed molecules are oriented in the same direction. It is worth noting that these transformations don't affect the orientation of the adsorbed molecules in the Monte Carlo program. The process of transformation is temporarily made after every Monte Carlo move when the order parameter is calculated. The susceptibility is constructed from fluctuations in the order parameter, *via*.

$$\chi = \left[ \langle \Phi^2 \rangle - \langle \Phi \rangle^2 \right] N / k_B T \quad (5)$$

To determine how often we should sample a set of available data ( $X$ ), in order to calculate quantities such as the average energy, heat capacity and order parameter, an examination of the autocorrelation function [142] of data  $\{X\}$  can be used,

$$C(\tau) = \frac{\langle X_i X_{i+\tau} \rangle - \langle X_i \rangle^2}{\langle X_i^2 \rangle - \langle X_i \rangle^2} \quad (6)$$

where  $X_i$  is the value of  $X$  at cycle  $i$ . The autocorrelation function for a number  $\tau$  of steps can be calculated. It has a value of one when the data are completely correlated ( $\tau = 0$ ) and decays exponentially to zero as  $\tau$  becomes large enough that the data became uncorrelated. We can then find the number of steps ( $\tau_e$ ) required for the autocorrelation function  $C(\tau)$  to decay to 0.367 ( $1/e$ ) of its value at  $\tau = 0$ . The data  $X$  can then be sampled every ( $\tau_e$ ) steps and still be uncorrelated. Phase transitions in real systems (in experiments) are considered to be in the thermodynamic limit and thus effectively infinite in size. In computer simulations, the sizes of adsorbed systems are finite and small compared to the sizes of the real systems. It has been proved that, in the finite systems used in computer simulations the phase transition temperature is shifted compared to the phase transition in the real systems. Hence, the transition temperature  $T_c(L)$  changes as the 2-d size  $L \times L$  of the system changes. This, plus the presence of large fluctuations due to finite size effects, makes it difficult to determine  $T_c(\infty)$ . In order to overcome these problems in determining the transition temperature for a system of infinite size, the Binder fourth order cumulant [26] [27]

$$U_L(T) = 1 - \frac{\langle \Phi^4 \rangle_L}{3 \langle \Phi^2 \rangle_L^2} \quad (7)$$

for several values of  $L$  can be used to locate the transition point for a system of infinite size. When  $U_L$  is plotted as a function of temperature for a several systems of different sizes, we will get a set of curves that intersect at the infinite size transition temperature where they are independent of the lattice size. The fourth order cumulant has two limiting values, for a completely ordered system  $U_L=2/3$ , while  $U_L=0$  when the system is completely disordered. As the system size is increased, the fourth order cumulant will have values close to these limits.

#### 4. Interaction potentials

To simulate the systems of interest of molecules on ionic surfaces, the potential energy functions must be constructed to calculate the total potential energy of the system to be simulated. The total potential consists of two parts, molecule-molecule and molecule-surface potentials. Our ability to reproduce known experiments data depends greatly on the parameters used to calculate the potential energy. Usually atom-atom/ion potentials with summation over a two body interactions is applied to describe the repulsion and dispersion interactions. In addition to these interactions, electrostatic contributions are also considered in similar way [1][29][28].

#### 4.1 Molecule-molecule interactions

Site-site model could be used to construct the intermolecular potential. The parameters governing the potentials are chosen so as to reproduce various experimental molecular

##### 4.1.1 Electrostatic interactions

The electrostatic interactions are mediated by point charges  $q_i$  and point dipoles  $\bar{\mu}_i$  that are distributed around the molecule (on the atomic sites) in such a way that the known multipole moments of the molecule can be reproduced through the following linear equations [92],

$$q_1 + q_2 = 0 \quad (8)$$

$$\mu = \sum_{i=1}^2 q_i r_i + \sum_{i=1}^2 \mu_i \quad (9)$$

$$\Theta = \sum_{i=1}^2 q_i r_i^2 + 2 \sum_{i=1}^2 \mu_i r_i \quad (10)$$

$$\Omega = \sum_{i=1}^2 q_i r_i^3 + 3 \sum_{i=1}^2 \mu_i r_i^2 \quad (11)$$

$$\Lambda = \sum_{i=1}^2 q_i r_i^4 + 4 \sum_{i=1}^2 \mu_i r_i^3 \quad (12)$$

where  $\mu$ ,  $\Theta$ ,  $\Omega$  and  $\Lambda$  are the molecular dipole, quadrupole, octupole and hexadecapole moments respectively. The  $r_i$  are the distances of the point charges  $q_i$  and point dipoles  $\mu_i$  at site  $i$  from the molecular center of mass. The values of the distributed point charges and dipoles at the atomic sites can be determined by using the above equations. The electrostatic interactions between an atom  $i$  (with charge  $q_i$  and dipole  $\mu_i$ ) of a molecule "m" with an atom  $j$  of another molecule "n" (with charge  $q_j$  and dipole  $\mu_j$ ) can be written as follows [96],

$$V_{ij}^{(elc)} = \frac{q_i q_j}{r_{ij}} - \frac{q_i \mu_j}{r_{ij}^3} (\hat{u}_j \cdot \bar{r}_{ij}) + \frac{q_j \mu_i}{r_{ij}^3} (\hat{u}_i \cdot \bar{r}_{ij}) + \frac{\mu_j \mu_i}{r_{ij}^5} \left[ (\hat{u}_i \cdot \hat{u}_j) r_{ij}^2 - 3 (\hat{u}_i \cdot \bar{r}_{ij}) (\hat{u}_j \cdot \bar{r}_{ij}) \right] \quad (13)$$

where  $\bar{r}_{ij}$  is the vector position of  $i^{\text{th}}$  atom of the  $n^{\text{th}}$  molecule with respect to the  $j^{\text{th}}$  atom of the  $m^{\text{th}}$  molecule. In other words,  $\bar{r}_{ij}$  points from atom  $j$  to atom  $i$ . And  $\hat{u}_i$  is an orientation unit vector of  $n^{\text{th}}$  molecule.

##### 4.1.2 Van der Waals interactions

Once again the atomic site model is usually used to model the van der Waals interaction (repulsion and dispersion interaction) between molecules so that, the molecular interaction may be expressed in terms of the atom-atom interactions by the modified Buckingham potential.

$$V_{ij}(r_{ij}) = A_{ij} \exp(-\eta_{ij} r_{ij}) - \frac{C_6^{ij}}{r_{ij}^6} - \frac{C_8^{ij}}{r_{ij}^8} \quad (14)$$

where the variable  $r_{ij}$  is the distance between atom sites  $i$  and  $j$  of different molecules. The Born-Mayer parameters,  $A_{ij}$  and  $\eta_{ij}$ , characterize the strength and the range of the repulsion respectively.  $C_6$  and  $C_8$  are dispersion constants that represent the strength of the leading terms in the dispersion series, these constants are well known for molecules and are easy to determine for atomic sites. In contrast, the repulsive parameters are difficult to obtain experimentally in most cases and hence the repulsive parameters can be adjusted so as to reproduce certain experimentally known quantities such as the crystal structure and cohesive energy of molecules on ionic surface. Known radii ( $a_i$ ) and softness ( $b_i$ ) parameters are used to construct the Born-Mayer parameters as follows [30],

$$\eta_{ij} = \frac{1}{b_i + b_j} \quad (15)$$

$$A_{ij} = (b_i + b_j) \exp\left(\frac{a_i + a_j}{b_i + b_j}\right) \quad (16)$$

In order to use the site-site model, the molecular dispersion interaction parameters, available in the literature for the molecules must be broken up into atomic based ones [32],

## 4.2 Molecule-surface interaction

The pairwise sum of two-body interactions (atom-ion interactions) could be used to model the electrostatic, dispersion and repulsion interactions between a molecule and the substrate. Because we are dealing with physisorbed systems where there is no noticeable reconstruction or distortion of the (001) ionic surface [33], it is reasonable to assume that the surface of the substrate is not perturbed by adsorbed molecules and hence has the same lattice constant as the bulk. Ions of the surface are considered to be periodic in two dimensions and regular (the substrate is regarded as a semi-infinite solid) in the third dimension

### 4.2.1 Electrostatic interactions

The electrostatic energy ( $V_{elc}^{m-s}$ ) of a single diatomic molecule on the ionic surface can have the following form

$$V_{elc}^{m-s} = \sum_{i=1}^2 \left( \psi(\vec{r}_i) q_i + \vec{E}(\vec{r}_i) \cdot \vec{\mu}_i - \frac{1}{2} \alpha_i^\perp E_\perp^2(\vec{r}_i) - \frac{1}{2} \alpha_i^\parallel E_\parallel^2(\vec{r}_i) \right) \quad (17)$$

Where  $q_i$ ,  $\vec{\mu}_i$  are point charges and point dipoles at the atomic sites of the molecule.  $\psi(\vec{r}_i)$  and  $\vec{E}(\vec{r}_i)$  are the electrostatic potential and electric field generated by the ionic crystal at position  $\vec{r}$ , the location of the interacting atom with respect to the origin of the coordinate system. The induction energy, which depends upon the "atomic" polarizabilities

(perpendicular ( $\alpha_i^\perp$ ) and parallel ( $\alpha_i^\parallel$ ) to the molecular axis), is included in our calculations.

The sum in the above equation is over the two atomic sites of the diatomic molecule. The electrostatic potential above the surface of a FCC ionic crystal is well known [34] and may be written as a two dimensional Fourier series whose leading term is [35]

$$\psi(\vec{r}_i) = -\frac{4e}{a} \left[ \frac{\exp\left(-\frac{2\pi z}{a}\right)}{1 + \exp\left(-\sqrt{2}\pi\right)} \right] \left( \cos\left(\frac{2\pi x}{a}\right) + \cos\left(\frac{2\pi y}{a}\right) \right) \quad (18)$$

where  $e$  is the absolute value of the electronic charge on an individual ion and  $a$  is the lattice constant of the surface mesh. The electric field at the crystal surface is readily calculated from the gradient of the electrostatic potential,

$$\vec{E}(\vec{r}) = -\vec{\nabla}\psi(\vec{r}) \quad (19)$$

#### 4.2.2 Van der Waals interactions

The Tang-Toennies potential is used to describe the repulsion and dispersion interaction of an atom of an adsorbed molecule with an ion of the substrate,

$$V_{ij}^{m-s}(r_{ij}) = A_{ij} \exp(-\eta_{ij}r_{ij}) - \sum_{n=3}^{\infty} f_{2n}(r_{ij}) \frac{C_{ij}^{2n}}{r_{ij}^{2n}} \quad (20)$$

where  $r_{ij}$  is the distance of atom  $i$  to ion  $j$  and  $C_6$ ,  $C_8$  and  $C_{10}$  are the atom-ion dispersion coefficients. The mathematical singularities at  $r=0$  are removed by the presence of the phenomenological damping functions

$$f_{2n}(r_{ij}) = 1 - \sum_{k=0}^{2n} \frac{\left(\eta_{ij}r_{ij}\right)^k}{k!} \exp(-\eta_{ij}r_{ij}) \quad (21)$$

The dispersion series is in principal infinite but some times for practical reasons is truncated at the  $k=5$  term, *i.e.* only the  $C_6$ ,  $C_8$  and  $C_{10}$  terms are included. To fully describe the potential it was necessary to estimate all interaction parameters  $A_{ij}$ ,  $\eta_{ij}$ ,  $C_6$ ,  $C_8$  and  $C_{10}$  for each of the atom-ion interactions.

##### 4.2.2.1 Dispersion parameters

Values of  $C_6$ ,  $C_8$  and  $C_{10}$  for the molecule-ion interactions can be estimated from combining rules derived [36] [37]. For example, the  $C_6$  are assumed to obey the following relation

$$C_6^{ij} = \frac{2C_6^{ii}C_6^{jj}\alpha^i\alpha^j}{C_6^{ii}(\alpha^j)^2 + C_6^{jj}(\alpha^i)^2} \quad (22)$$

where the index  $i$  refers to the molecule and  $j$  refers to either the positive or negative ion in the substrate, and  $\alpha$  is average polarizability. The values of  $C_8$  can be found using the following relations [36]:

$$C_8^{ij} = C^{ij}(1,2) + C^{ij}(2,1) \quad (23)$$

Where,

$$C^{ij}(1,2) = \frac{15}{4} \left[ \frac{\alpha_1^i \alpha_2^j Y_1^i Y_2^j}{Y_1^i + Y_2^j} \right] \quad (24)$$

$$Y_1^i = \frac{4}{3} \frac{C_6^i}{(\alpha_1^i)^2} \quad (25)$$

$$Y_2^j = \frac{2C_8^i Y_1^i}{15\alpha_1^i \alpha_2^j - 2C_8^i} \quad (26)$$

$$C_6^{ij} = \frac{2C_6^{ii} C_6^{jj} \alpha^i \alpha^j}{C_6^{ii} (\alpha^j)^2 + C_6^{jj} (\alpha^i)^2} \quad (27)$$

The values of the  $C_{10}$  coefficients for each of the molecule-ion pairs were estimated using the approximate relation [38]

$$C_{10} = 49(C_8)^2 / 40C_6 \quad (28)$$

These values of the molecule-ion dispersion coefficients  $C_6$ ,  $C_8$ , and  $C_{10}$  were used to calculate the atom-ion dispersion coefficients

$$C_6^{\text{atom-ion}} = \left( \frac{\alpha_{\text{atom}}}{\alpha_{\text{molecule}}} \right) C_6^{\text{molecule-ion}} \quad (29)$$

#### 4.2.2.1 Repulsion parameters

The method used to obtain Born-Mayer parameters for the atom-ion interactions are identical to that used in Ref. 29. The Born-Mayer parameters for atom-ion interactions are estimated by using the combining rules of Gilbert [39] [40] and Smith [41]

$$A_{ij} = \left[ \frac{\eta_{ii} + \eta_{jj}}{2\eta_{ii}\eta_{jj}} \right] (A_{ii}\eta_{ii})^c (A_{jj}\eta_{jj})^d \quad (30)$$

Where,

$$\eta_{ij} = \frac{2\eta_{ii}\eta_{jj}}{(\eta_{ii} + \eta_{jj})} \quad (31)$$

$$c = \frac{\eta_{jj}}{\eta_{ii} + \eta_{jj}} \quad (32)$$

$$d = \frac{\eta_{ii}}{\eta_{ii} + \eta_{jj}} \quad (33)$$

## 5. Monte Carlo results

A great deal of understanding of the fundamental processes in surface science comes through the use various experimental and computational studies of adsorbed layers on various solid surfaces, e.g. the theory of interactions of gases with solid surfaces [1][2], orientational ordering of adsorbed layers, e.g. diatomic molecules on graphite [3][4][5], thin films on solid surfaces [6], surface diffusion [7], surface aligned photochemistry [8][9], phase transitions and critical phenomena, e.g. 4He on graphite [10][11], 4He on Kr-preplated graphite [12], and the structures and dynamics of molecules on ionic surfaces [13].

Parallel to the experimental techniques such as low energy electron diffraction (LEED)[14], helium atom scattering (HAS)[15][16], polarization infrared spectroscopy (PIRS)[17] and Calorimetric, computer simulations have been accepted as a useful tool in determining additional details of the structures of adsorbed molecules. Thus they help in understanding physics at surfaces [20][21], and have been used to substantiate and interpret experimental results [22-26]. Furthermore, computer simulations can be used to predict additional results and guide experiments to perform more experimental work [5][27]. The most widely known techniques in computer simulation are the Metropolis Monte Carlo (MC) [26-29] and Molecular Dynamics (MD) methods [21].

### 5.1 N<sub>2</sub> and CO on NaCl(001) and LiF(001)

Monte Carlo simulations have been used to study the structures and phase transitions of CO/NaCl(001), CO/LiF(001) and N<sub>2</sub>/NaCl(001) systems [15][16][18]. Through the use of Monte Carlo simulations, these systems have been identified as falling into the class of phase transition whose critical exponents are nonuniversal.

What makes the critical exponents interesting is the idea of universality. The critical exponents are found to be independent of the details of the interatomic interactions. According to the idea of universality, the critical exponents of all systems that exhibit a continuous phase transition near the critical temperature can be grouped into a small number of universality classes. Within each universality class, the critical behavior is remarkably similar. In Table 1 the universality classes, which are related to the order-disorder phase transitions in two dimensions are presented.

In two dimensional systems there is a symmetry class whose critical exponents have nonuniversal values [43-46]. The values of the exponents in this class depend on the strengths of an anisotropic external potential  $h_4$ . In CO/NaCl(001), CO/LiF(001) and N<sub>2</sub>/NaCl(001) systems this anisotropy is provided by the substrate and perhaps the molecule-molecule interactions. As shown in Fig.3, for infinite anisotropy the Ising exponents are recovered whereas in the limit of zero anisotropy Kosterlitz-Thouless (K-T) behaviour occurs. For example, the critical exponent  $\beta$ , has a value of 0.125 for the Ising model (infinite anisotropy) and will increase towards infinity as K-T behaviour is

approached at zero anisotropy. The critical exponents  $\alpha$ ,  $\beta$ , and  $\gamma$  are associated, respectively, with the critical behaviour of the heat capacity  $C_v$ , order parameter  $\Phi$ , and susceptibility  $\chi$  as follows:  $C_v \sim A^\pm t^{-\alpha}$ ,  $\Phi \sim B^\pm t^{-\beta}$ , and  $\chi \sim C^\pm t^{-\gamma}$ , where the reduced temperature is defined as  $t = |(T_c - T)/T_c|$  and the amplitudes  $A^\pm$ ,  $B^\pm$ , and  $C^\pm$  carry a positive (negative) superscript for the temperatures above (below)  $T_c$ . With values of  $\alpha$ ,  $\beta$  and  $\gamma$  in hand it is possible to check the validity of Rushbrooke's relation,  $\alpha + 2\beta + \gamma \geq 2$ .

Universality class Exponent	Ising	XY with cubic anisotropy	3-state Potts	4-state Potts
$\alpha$	$O(\log)$	Non universal	1/3	2/3
$\beta$	1/8	Non universal	1/9	1/12
$\gamma$	7/4	Non universal	13/9	7/6

Table 1.1 Universality classes and its critical exponents.

Monte Carlo (MC) simulations have provided details of the ordered monolayer structure and successfully reproduced the transition to the disordered state at temperatures around 30–35 K in the CO/NaCl(001) system [15]. It was argued that these phase transitions are of interest because they fall into the universality class of the XY model with cubic anisotropy and hence should have nonuniversal critical exponents, i.e., their values depend on the relative strengths of an anisotropic external potential provided by the substrate and the molecule molecule interactions.

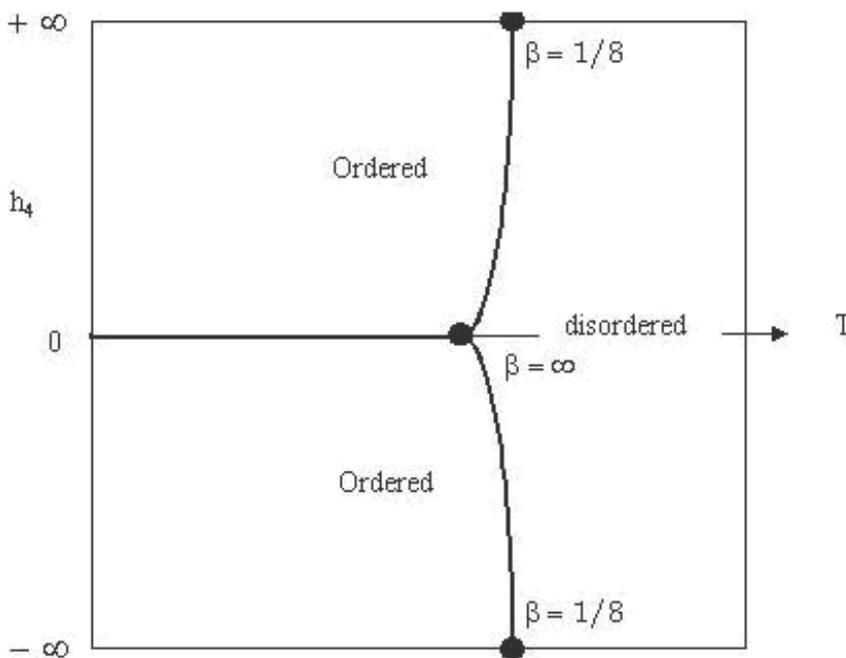


Fig. 3. Phase diagram for the XY model in the  $h_4 - T$  plane shows how the critical exponent  $\beta$  varies with anisotropy strength  $h_4$  (Ref. [43]).

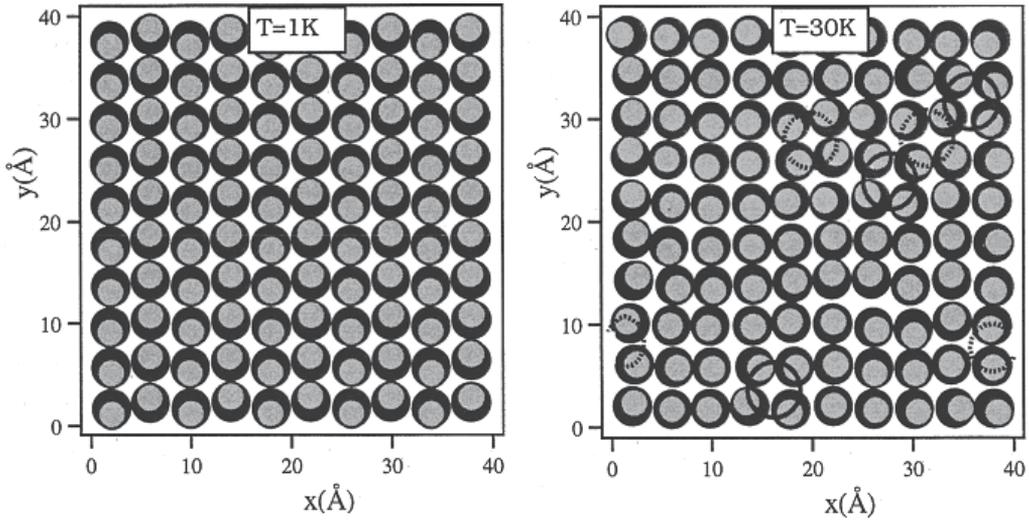


Fig. 4. An overview of a typical configuration of 100  $N_2$  molecules at 1 K and 30 K. The nitrogen atoms closest to the surface are shown as black and the upper nitrogen atoms as gray. Clockwise (counterclockwise) vortices are denoted by the solid (dashed) circles. (this figure taken from Ref [18])

Metropolis Monte Carlo (canonical ensemble) simulations of the  $N_2$  /NaCl(001) [18] system is predict that at low temperatures a monolayer of nitrogen molecules forms an ordered  $p(2 \times 1)$  structure which, upon heating past 25 K, undergoes an order-disorder phase transition as shown in Fig. 4. In the disordered phase the long-range orientational order among molecules is destroyed although residual shortrange order persists in the form of small ordered domains and pairs of counter rotating vortices. The destruction of orientational order is further shown in Fig. 5, where the azimuthal angle distributions are plotted for several temperatures. The distributions are sharply peaked at  $\pm 90^\circ$  at low temperatures and broaden as the temperature increases. Around 25 K the distribution becomes uniform indicating a loss of long-range azimuthal order and signaling a transition to a disordered phase. However, minor residual peaks at  $\varphi = \pm 90^\circ$ , are still observed at  $T=30$  K. As one might expect, this overall behavior is similar to that of the CO/NaCl system with differences occurring in the details, such as the values of the transition temperature and tilt angle. The heat capacity of the  $N_2$  /NaCl(001) was found to have a maximum at 25.0 K and exhibit a logarithmic type divergence that can be expressed as

$$C-/R = -0.256 \ln(t) - 0.158 + 2.5 \text{ for } T < T_c, \quad (34)$$

$$C+/R = -0.250 \ln(t) - 0.200 + 2.5 \text{ for } T > T_c, \quad (35)$$

where the background heat capacity of  $2.5R$ , due to the presence of three translational and two rotational modes, is shown explicitly. The slopes are close to those reported previously. The heat capacity data was also analyzed as a power law divergence and a value of  $\alpha = 0.076 \pm 0.010$  was found [18].

The adsorption of  $N_2$  on the LiF(0 0 1) surface is studied by canonical Monte Carlo (CMC) computer simulation [19]. As shown in Fig. 6, these studies predicted that  $N_2$  forms an ordered structure where the molecules are arranged in a unit cell of  $p(2\sqrt{2} \times \sqrt{2})R45^\circ$

symmetry at temperatures below 23 K with 50% coverage. As shown in Fig. 7, the nitrogen molecules are tilted by  $53^\circ$  from the surface normal and have the same azimuthal orientation along diagonals, with diagonals alternating their orientation see Fig. 7. Beyond 23 K, the molecules become azimuthally disordered but with residual short-range order. No change in the position of the peak of the polar (tilt) angle distribution was observed above the transition temperature. This transition is purely of the order-disorder type.

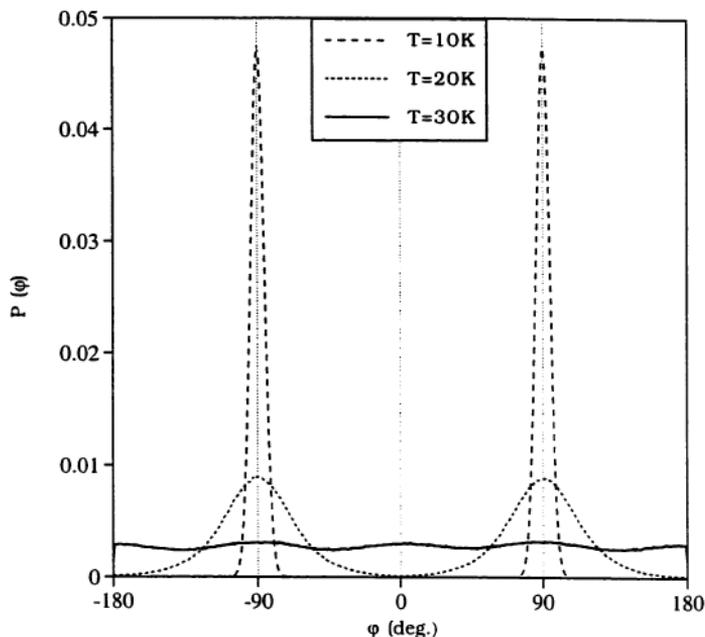


Fig. 5. The azimuthal angle ( $\varphi$ ) distribution is plotted for temperatures  $T=10, 20, 30$  K. At 1K the distributions are symmetric and centered on the  $\theta \sim 0^\circ, 31^\circ$ . As the temperature increases this peak decreases in height and broadens in width. The peak centered on  $\theta \sim 31^\circ$  at 1K shifts below to  $\theta \sim 28^\circ$  at 40 K.

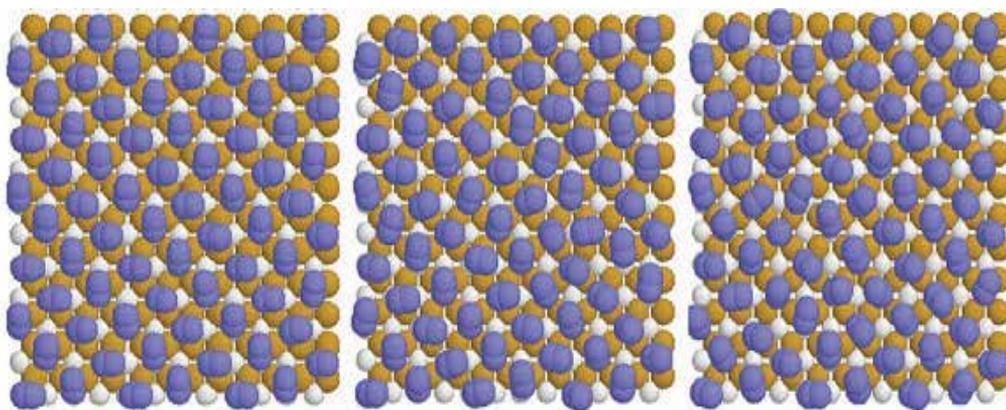


Fig. 6. The final configurations of a monolayer of  $N_2$  on LiF(001) surface at 1 K (left), 20 K (middle) and 25 K (right). The monolayer at 1K forms an ordered  $p(2\sqrt{2} \times \sqrt{2})R45^\circ$  structure

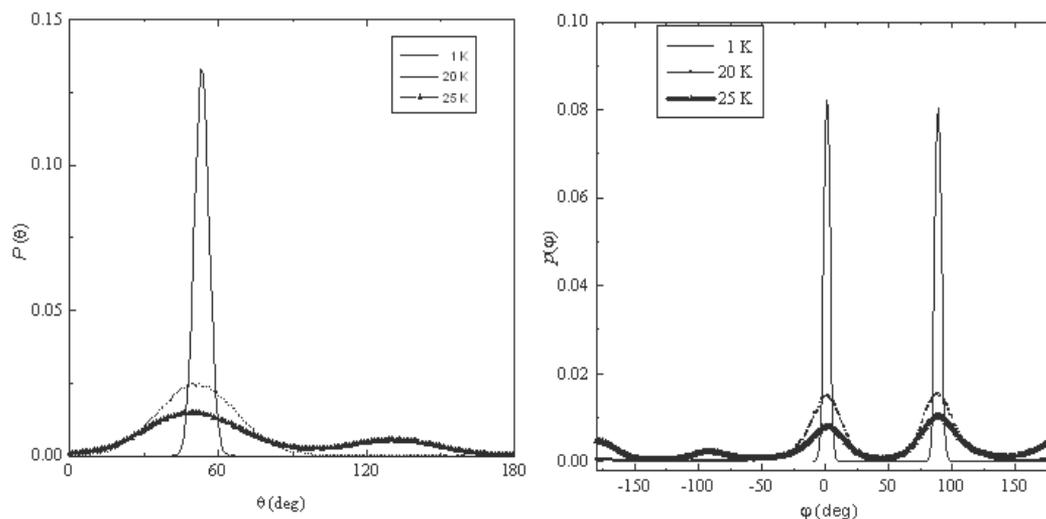


Fig. 7. Polar angle (left) distribution is plotted for temperature  $T = 1$  K, 20 K and 25 K. At 1 K the distribution is symmetric and centered at  $53^\circ$ . Azimuthal angle ( $\varphi$ ) distributions (right) of the  $N_2$  molecules adsorbed on LiF surface at 1 K, 20 K and 25 K show the progress of the transition from an ordered to a disordered phase.

## 5.2 CO on MgO(001)

Monte Carlo simulations of CO [17] show that below 41 K the CO molecules form a  $c(4 \times 2)$  structure with six molecules per unit cell distributed into two kinds of adsorption sites as shown in Fig. 8: a perpendicular site and a tilted site (polar angle of  $31^\circ$ ). Both sites are localized near  $Mg^{2+}$  ions. The occupancy of perpendicular sites to tilted sites occurs in the ratio of 1:2. At 41 K the  $c(4 \times 2)$  phase undergoes a phase transition into a less dense, disordered phase accompanied by the expulsion of some molecules to form a partial second layer. The density of the remaining disordered layer is the same as for a  $p(3 \times 2)$  phase and portions of the disordered layer show regions of short range ordering with either the  $c(4 \times 2)$  or  $p(3 \times 2)$  structures. The  $p(3 \times 2)$  phase contains four molecules per unit cell and also consists of perpendicular and tilted sites, but in the ratio of 1:1. This structure was found to be stable up to 50 K after which the expulsion of some molecules and disordering of the layer occurred. A model to test the relative stability of these two phases by examining the difference in Gibbs free energy is constructed and shows that below 41 K the  $c(4 \times 2)$  phase is the most stable but above 41 K the  $p(3 \times 2)$  phase is the most stable. However, at low pressures the model suggests that the  $p(3 \times 2)$  phase will not be observed and the layer will instead transform from the  $c(4 \times 2)$  phase to a disordered phase at 41 K. This result reconciles the findings of low-energy electron diffraction (LEED) experiments [ $p(3 \times 2)$  phase observed] with those of helium atom scattering (HAS) and polarization infrared spectroscopy (PIRS) experiments (disordered phase observed). It is proposed that the  $c(4 \times 2) \rightarrow p(3 \times 2)$  transition is part of an infinite sequence of transitions involving  $(n \times 2)$ -type structures which, under suitable conditions of temperature and pressure, constitutes an example of the devil's staircase phenomenon. Such a phenomenon has been suggested by previous LEED experiments.

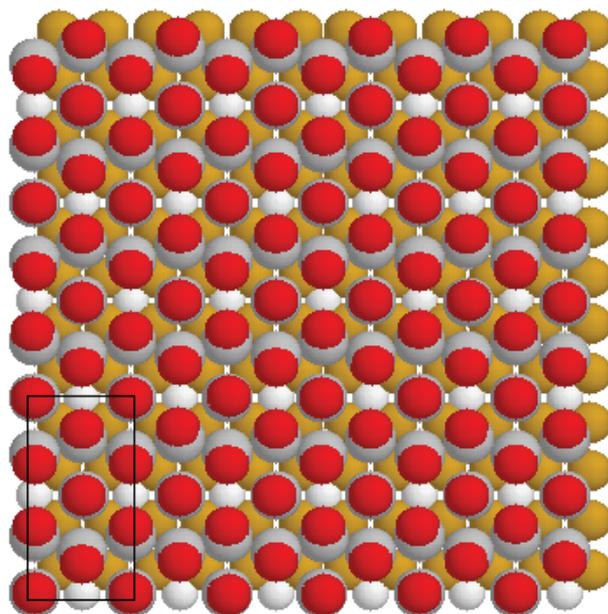


Fig. 8. A top view of the MgO(100) surface covered with 108 CO molecules at 1 K. The carbon atoms are shown as white and the oxygen atoms as red. The small white ball represents a Mg ion and orange ball represents a O ion. Note that the origin is centered on a O ion. The resulting  $c(4 \times 2)$  unit cell is shown (solid lines).

### 5.3 H<sub>2</sub> and D<sub>2</sub> on NaCl(001), LiF(001) and MgO(001)

A Monte Carlo simulation is used to study the H<sub>2</sub> and D<sub>2</sub> molecules adsorbed on a NaCl(001), LiF(001) and MgO(001) surfaces [20-23]. In the case of H<sub>2</sub> on NaCl(001), (Fig. 10) H<sub>2</sub> forms a commensurate  $c(2 \times 2)$  structure where the hydrogen molecules sit flat on top of the cationic Na<sup>+</sup> sites. The unit cell was found to have four molecules, where pairs of neighboring molecules are aligned perpendicular to each other in a "T" configuration. This structure is in agreement with the experimental results in terms of coverage and stability, but disagrees in terms of symmetry since the PIRS-ATR and HAS experimental results show a  $(1 \times 1)$  structure. To solve this problem, the rotational motion of H<sub>2</sub> molecules has been studied using perturbation theory and it is found that quantum effects will azimuthally delocalize the orientation of the molecular axis of H<sub>2</sub>. Thus, the  $c(2 \times 2)$  structure becomes a  $(1 \times 1)$  structure. These simulations also show that a second layer is possible, where all hydrogen molecules adsorb over the anionic sites in a unit cell of  $p(2 \times 1)$  symmetry. Perturbation Theory calculations show that *p*-H<sub>2</sub> and *o*-H<sub>2</sub> ( $J=1, m=\pm 1$ ) prefer to sit on the top of Na<sup>+</sup> site, while *o*-H<sub>2</sub> ( $J=1, m=0$ ) prefers to locate over the Cl<sup>-</sup> site. Monte Carlo (MC) simulations of D<sub>2</sub> molecules on the MgO(001) surface are reported and show that a series of interesting structures form with increasing coverage, viz.  $p(2 \times 2) \rightarrow p(4 \times 2) \rightarrow p(6 \times 2)$ , with coverages  $\theta = 0.5, 0.75,$  and  $0.83$  respectively, and are stable up to 13 K. The  $p(2 \times 2)$  structures contain two D<sub>2</sub> molecules per unit cell, with each molecule lying parallel to the plane of the surface ( $\theta = 90^\circ$ ) directly above every other Mg<sup>2+</sup> site. The molecules adopt a "T" configuration with respect to their nearest neighbors. The  $p(4 \times 2)$  and  $p(6 \times 2)$  structures,

have two kinds of adsorption sites: a parallel site, as in the case of  $p(2 \times 2)$ , and a tilted site, where the  $D_2$  molecules sit between cationic and anionic sites with the molecular axis directed towards the anionic site, with  $\theta \approx 60^\circ$ . These structures are consistent with recent Neutron Scattering results in terms of coverage and stability, but disagree in terms of symmetry; the neutron scattering work found "c" type structures whereas the MC simulations (without quantum considerations) yield a "p" type structures. To reconcile the results of the simulations and experiments, the quantum mechanical rotational motion of the adsorbed  $D_2$  molecules was studied using perturbation theory. These calculations show that the adsorbed  $D_2$  molecules are azimuthally delocalized and hence the structures are indeed "c" type rather than "p" type.

Monte Carlo (MC) simulations has been preformed for  $H_2$  on LiF(001). MC simulations predict that  $H_2$  molecules form a series of interesting structures,  $p(2 \times 2) \rightarrow p(8 \times 2) \rightarrow p(4 \times 2)$  with coverages  $\Theta=0.5, 0.625$  and  $0.75$  respectively, that are stable up to 8 K (see Fig. 11). These structures are consistent with recent Helium Atom Scattering results (the  $p(4 \times 2)$  is not observed) in terms of coverage and stability, but disagree in terms of symmetry. The HAS work found "c" type structures whereas the Metropolis MC simulations yield a "p" type structures. To reconcile the results of the simulations and experiments, the rotational motion of the adsorbed  $H_2$  molecules was studied using perturbation theory. These calculations show that the adsorbed  $H_2$  molecules are azimuthally delocalized and hence the structures are indeed c-type. Our calculations also indicate that p- $H_2$  and helicoptering o- $H_2$  prefer cationic sites, while cartwheeling o- $H_2$  prefers anionic sites.

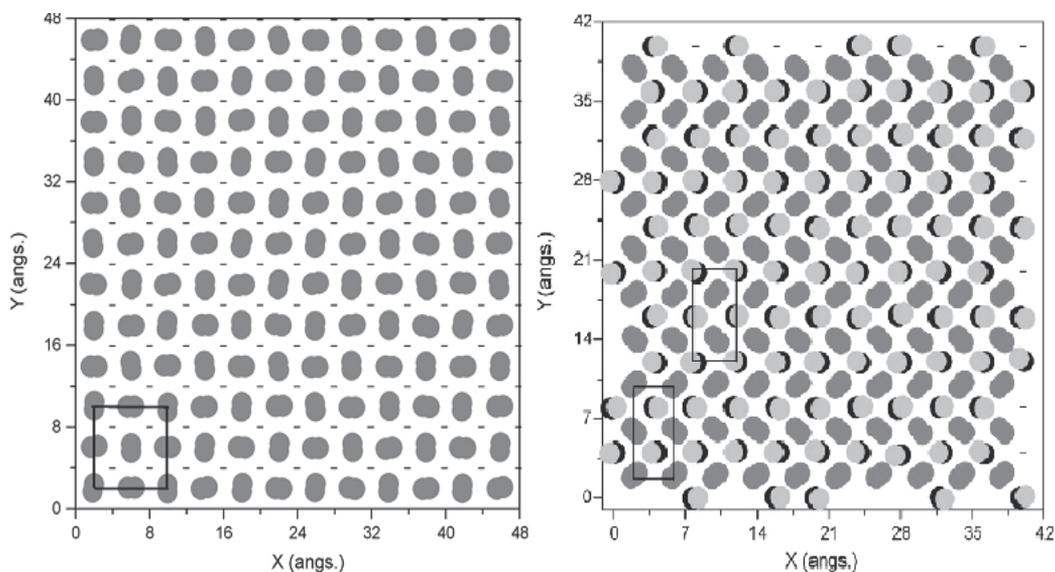


Fig. 11. The LiF(001) surface covered with  $H_2$  molecules at 1 K. The (blue) symbol represents a  $Li^+$  ion and the (yellow) symbol represents a  $F^-$  ion. The hydrogen atoms of molecules are shown in white color.  $p(8 \times 2)$  structure shown in the right image,  $p(4 \times 2)$  structure shown in the middle image and  $p(2 \times 2)$  structure shown in the left image.

Monte Carlo simulations show that  $D_2$  adopt a sequence of  $c(n \times 2)$  structures. The  $c(2 \times 2)$  structure consists of an array of molecules covering every other  $Mg^{2+}$  site of the surface in a checkerboard pattern, with quantum mechanical delocalisation of the molecular axes eliminating azimuthal differences between molecules. Specifically, the *ortho* and helicoptering *para* states are allowed to adsorb here. The  $c(4 \times 2)$  structure consists of two kinds of adsorption sites. One third of the molecules adsorb directly over  $Mg^{2+}$  ions with a preference for a horizontal orientation for the molecular axes (*ortho* or helicoptering *para*-states), while the remaining two thirds adsorb near, but are offset from, the  $O^{2-}$  ions with orientations that prefer a tilt from the surface normal. In terms of rotational states these are thought to be a mix of cartwheeling and helicoptering *para*-states or possibly skewed *ortho*-states. The tilted molecules sit  $0.5 \text{ \AA}$  further from the surface than the horizontal molecules. The  $c(6 \times 2)$  structure is an extension of the  $c(4 \times 2)$  structure with only  $1/5^{\text{th}}$  of the molecules adopting a horizontal orientation; the rest are tilted near  $O^{2-}$  ions.

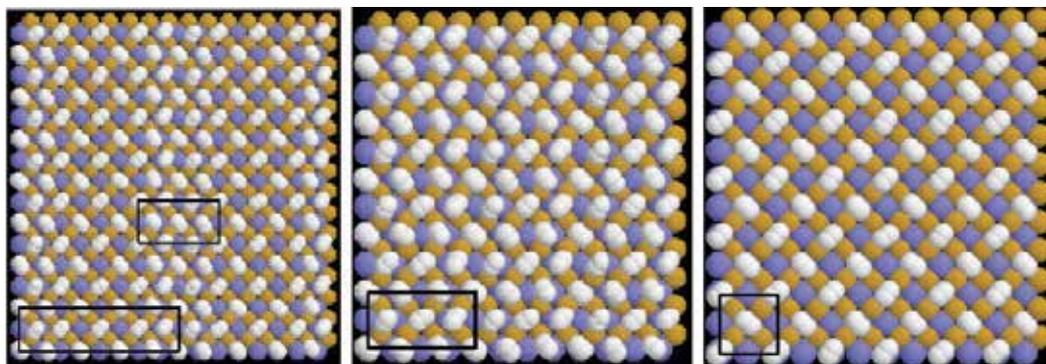


Fig. 10. The LiF(001) surface covered with  $H_2$  molecules at 1 K. The (blue) symbol represents a  $Li^+$  ion and the (yellow) symbol represents a F-ion. The hydrogen atoms of molecules are shown in white color.  $p(8 \times 2)$  structure shown in the right image,  $p(4 \times 2)$  structure shown in the middle image and  $p(2 \times 2)$  structure shown in the left image.

## 6. Grand Canonical Monte Carlo (GCMC) simulation for $CO_2$ on $MgO(001)$

The adsorption isotherms of  $CO_2$  on  $MgO$  is obtained using Grand Canonical Monte Carlo (GCMC) simulations and compared with experiment, as well as to explore the possible formation of monolayers of different densities [42]. The Canonical Monte Carlo (GCMC) refers to a simulation when the system at fixed temperature  $T$ , volume  $V$  and chemical potential  $\delta$ . The chemical potential of the gas above the surface layer of the adsorbed molecules is in equilibrium with the adsorbed molecules:

$$\delta_{\text{surface}} = \delta_{\text{gas}} \quad (36)$$

If the gas is considered to be ideal, the relation between the chemical potential and the pressure of the adsorbed molecules is

$$\delta_{\text{molecules}} = k_B T \ln(P \omega^3 / k_B T) \quad (37)$$

Where  $\varpi$  is the thermal de Broglie wavelength and  $k_B$  is the Boltzmann constant. The GCMC simulations are actually performed at constant  $B$ ,  $V$ , and  $T$ , where  $B$  is the so-called Adams parameter defined as

$$\delta = k_B T B - k_B T \ln(\varpi^3 / V) \quad (38)$$

The required relation for GCMC simulations is given as

$$B = \ln(P\varpi^3 / k_B T) \quad (39)$$

The formation of a high density monolayer of CO<sub>2</sub> on MgO(001) has been successfully simulated.

## 7. Conclusion

History and basics of Monte Carlo methods are discussed in the beginning of this chapter. Then we give methods for describing the coordinate system and potentials that govern the interaction of adsorbed molecules with surfaces. The use of Monte Carlo methods to test the modern theory of phase transition in real systems have been explained. Statistical techniques to analyze the data obtained from simulations have been discussed. Applications of Monte Carlo simulation of the real physical systems is discussed in details: e.g. N<sub>2</sub>, H<sub>2</sub> and D<sub>2</sub> adsorbed on NaCl(001), H<sub>2</sub>, D<sub>2</sub> and N<sub>2</sub> adsorbed on LiF(001), CO, CO<sub>2</sub>, H<sub>2</sub> and D<sub>2</sub> adsorbed on MgO(001).

## 8. References

- [1] M.P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Clarendon Press, Oxford, 1987.
- [2] M.P. Allen and D. J. Tildesley, *Computer Simulation in Chemical Physics*, Kluwer Academic Publishers, Dordrecht (1993).
- [3] M. E. J. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics*, Clarendon Press, Oxford (1999).
- [4] Kalos and Whitlock. *Monte Carlo Methods, Volume I: Basics*. John Wiley & Sons, 1986.
- [5] Hammersley and Handscomb. *Monte Carlo Methods*. John Wiley & Sons, 1965.
- [6] Jerome Spanier and Ely M. Gelbard. *Monte Carlo principles and neutron transport problems*. Reading, Mass., Addison-Wesley Pub. Co, 1969.
- [7] Christian P. Robert and George Casella. *Monte Carlo Statistical Methods*. Springer-Verlag, 2nd edition, 2004.
- [8] A. Hall. On an experimental determination of Pi. *Messeng. Math.*, 2:113–114, 1873.
- [9] N. Metropolis and S. Ulam. The Monte Carlo method. *Journal of the American Statistical Association*, 44:335–341, 1949.
- [10] D. Nicholson and N. G. Parsonage, *Computer Simulation and Statistical Mechanics of adsorption*, Academic Press, New York (1982).
- [10] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* 21,1078(1953).

- [11] J. J. Binney, N. J. Dowrick, A. J. Fisher and M. E. J. Newman, *The Theory of Critical Phenomena an Introduction to the Renormalization Group*, Oxford University Press (1992).
- [12] K. Binder, D. W. Hermann, *Monte Carlo Simulations in Statistical Physics an Introduction*, Springer-Verlag, Berlin (1992).
- [13] Polanyi, Williams and O'Shea, *J. Chem. Phys.* 94,978 (1991)
- [14] W. Hu, M.-A. Saberi, A. Jakalian, and D. B. Jack, *J. Chem. Phys.* 106, 2547 (1997).
- [15] N.-T. Vu, A. Jakalian, and D. B. Jack, *J. Chem. Phys.* 106, 2551 (1997).
- [16] N. -T. Vu and D. B. Jack, *J. Chem. Phys.* 108, 5653 (1998).
- [17] A. K. Sallabi and D. B. Jack, *J. Chem. Phys.* 112, 5133 (2000).
- [18] A. K. Sallabi and D. B. Jack. *Phys. Rev. B* 62, R4841 (2000).
- [19] A. K. Sallabi ,J. N. Dawoud , and D. B. Jack, *Applied Surface Science* 256,2974(2010).
- [20] J. N. Dawoud , A. K. Sallabi and D. B. Jack, *Applied Surface Science* 254, 7807(2008)
- [21] J. N. Dawoud, A. K. Sallabi, I. I. Fasfous, D. B. Jack, *Journal of Surface Science and Nanotechnology* 7,207 (2009).
- [22] J. N. Dawoud , A. K. Sallabi and D. B. Jack, *Surface Science* 601, 3731(2007).
- [23] Jamal Dawouda, Abdulwahab Sallabib, Ismail Fasfousa, and David Jack, *Jordan Journal of Chemistry.* 3,269( 2008).
- [24] W. Hu, M.Sc. Thesis, Concordia University (1997).
- [25] A. B. MacIsaac, J. P. Whithead, K. De'Bell, and P. H. Poole, *Phys. Rev. Lett.* 77, 739 (1996).
- [26] K. Binder, *Annu. Rev. Phys. Chem.* 37, 401 (1992).
- [27] K. Binder, *Phys. Rev. Lett.* 47, 693 (1981).
- [28] J. C. Polanyi and J. Williams, *J. Chem. Phys.* 94, 978 (1991).
- [29] V. J. Barcaly, D. B. Jack, J. C. Polanyi, and Y. Zeiri, *J. Chem. Phys.* 97, 9458 (1992).
- [30] M. A. Sabri, M. Sc. Thesis, Concordia University (1996).
- [31] M. Karplus and R. N. Porter, *Atoms and Molecules*(Benjamin/Cummings, Menlo Park, 1970).
- [32] F. Mulder, G. F. Thomas, and W. J. Meath, *Mol. Phys.* 41, 249 (1980).
- [33] A. W. Meredith and A. J. Stone, *J. Chem. Phys.* 104, 3058 (1996).
- [34] J. E. Lennard-Jones and B. M. Dent, *Trans. Farad. Soc.* 24, 92 (1928).
- [35] W. A. Steele, *The Interaction of Gases with Solid Surfaces* (Pergamon press, Oxford,1974).
- [36] T. Tang and J. P. Toennies, *Z. Phys. D*, 1, 91 (1986).
- [37] Habitz, P., Tang, K. T., Toennies, J. P., *Chem. Phys. Lett.* 85, 461 (1982).
- [38] C. Douketis, G. Scoles, S. Marchetti, M. Zen, and A. Thakkar, *J. Chem. Phys.* 76, 3057 (1982).
- [39] T. L. Gilbert, *J. Chem. Phys.* 49, 2640 (1968).
- [40] T. L. Gilbert, O. C. Simpson, and M. A. Williamson, *ibid* 63, 4061 (1975).
- [41] F. T. Smith, *Phys. Rev. A* 5, 1708 (1972).
- [42] Christopher D. Daub, G. N. Patey, D. B. Jack and A. K. Sallabi, "Monte Carlo simulations of the adsorption of CO<sub>2</sub> on the MgO(100) surface" *J. Chem. Phys.*124, 114706 (2006).
- [43] L. D. Roelofs and P. J. Estrup, *Surf. Sci.* 125, 51 (1983).

- [44] J. M. Yeomans, *Statistical Mechanics of Phase Transitions*, Clarendon Press, Oxford (1992).
- [45] J. V. José, L. P. Kadanoff, S. Kirkpatrick, and D. R. Nelson, *Phys. Rev. B* 16, 1217 (1977).
- [46] G. Y. Hu and S. C. Ying, *Physica* 140A, 585 (1987).



*Edited by Shaul Mordechai*

In this book, Applications of Monte Carlo Method in Science and Engineering, we further expose the broad range of applications of Monte Carlo simulation in the fields of Quantum Physics, Statistical Physics, Reliability, Medical Physics, Polycrystalline Materials, Ising Model, Chemistry, Agriculture, Food Processing, X-ray Imaging, Electron Dynamics in Doped Semiconductors, Metallurgy, Remote Sensing and much more diverse topics. The book chapters included in this volume clearly reflect the current scientific importance of Monte Carlo techniques in various fields of research.

Photo by litt1ehenrabi / iStock

**IntechOpen**

