IntechOpen

Recent Advances in Thermo and Fluid Dynamics

Edited by Mofid Gorji-Bandpy

RECENT ADVANCES IN THERMO AND FLUID DYNAMICS

Edited by **Mofid Gorji-Bandpy**

Recent Advances in Thermo and Fluid Dynamics
http://dx.doi.org/10.5772/59835
Edited by Mofid Gorji-Bandpy

Contributors

Vesna Dragicevic, Andras Szasz, Gyula Vincze, Alexandru Parvan, Shinya Shimokawa, Tomokazu Murakami, Akiyuki Ukai, Kouta Nakase, Hiroyoshi Kohno, Akira Mizutani, Bohdan Hejna, Riza Erdem, Mofid Gorji-Bandpy, Hossein Yahyazadeh, Delfino Ladino-Luna, Sohrab Rohani, Jack Denur

Notice

First published in Croatia, 2015 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

Recent Advances in Thermo and Fluid Dynamics
Edited by Mofid Gorji-Bandpy

p. cm.

ISBN 978-953-51-2239-5

eBook (PDF) ISBN 978-953-51-6650-4

Meet the editor

Dr. Mofid Gorji-Bandpy received his MS degree in Mechanical Engineering from the Faculty of Engineering, University of Tehran, Iran, in 1978. In 1990, he obtained his PhD degree in Hydraulic Engineering from the School of Engineering, University of Wales College of Cardiff (UWCC), UK. At present, he is a Full Professor in the Department of Mechanical Engineering at the Babol Noshirvani University of Technology, Babol, Iran. He is also a Visiting Professor in the Department of Mechanical and Industrial Engineering at the University of Toronto, Canada. He has visited more than 40 countries of the world. His major interests are in advanced methods of energy conversion systems, turbomachinery, fluid mechanics, water distribution network, and the solutions of both energy and environmental problems. He has published more than two hundred papers in these fields. He has also several refereed publications in different fields of mechanical engineering, applied mathematics, and aerodynamics. In addition, he has been an invited editor and reviewer for many international scientific journals.

Contents

Preface

The topic of thermodynamics is taught in physics and chemistry courses as part of the regular curriculum. Concepts of thermodynamics are used to solve engineering problems. Engineers use thermodynamics to calculate the fuel efficiency of engines and to find ways to make more efficient systems, be they rockets, refineries, or nuclear reactors. One aspect of "engineering" in the title is that a lot of the data used is empirical (e.g., steam tables), since you won't find clean algebraic equations of state for many common working substances. Thermodynamics is the science that deals with the transfer of heat and work. Engineering thermodynamics develops the theory and techniques required to use empirical thermodynamic data effectively. However, with the advent of computers, most of these techniques are transparent to the engineer, and instead of looking up data in tables, computer applications can be queried to retrieve the required values and use them in calculations. There are even applications tailored to specific areas which give answers for common design situations. But a thorough understanding will only come with the knowledge of underlying principles, and the ability to judge the limitations of empirical data is perhaps the most important gain from such knowledge.

Thermodynamics is the study of the relationships between HEAT (thermos) and WORK (dynamics). Thus, it deals with energy interactions in physical systems. Classical thermodynamics is based on the four laws of thermodynamics, called the zeroth, first, second, and third laws, respectively. The laws of thermodynamics are empirical, i.e., they are deduced from experience and supported by a large body of experimental evidence. These laws had a deep influence on the development of physics and chemistry.

In the past, historians considered thermodynamics as a science that is isolated, but in recent years, scientists have incorporated more friendly approaches to it and have demonstrated a wide range of applications of thermodynamics.

The book aims to present novel ideas that are crossing traditional disciplinary boundaries and introducing a wide spectrum of viewpoints and approaches in applied thermodynamics of the third millennium. The book will be of interest to those working in the fields of propulsion systems, power generation systems, chemical industry, quantum systems, refrigeration, fluid flow, combustion, and other phenomena. It can also be used in postgraduate courses for students and as a reference book, as it is written in a language pleasing to the readers.

Mofid Gorji-Bandpy
Department of Mechanical Engineering
Babol Noshirvani University of Technology
Iran

Chapter 1

Prediction of Solubility of Active Pharmaceutical Ingredients in Single Solvents and Their Mixtures — Solvent Screening

Ehsan Sheikholeslamzadeh and Sohrab Rohani

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/60982

Abstract

In this chapter, the applicability of two predictive activity coefficient-based models will be examined. The experimental data from five different types of VLE (vapor-liquid equilibrium) and VLLE (vapor-liquid-liquid equilibrium) systems that are common in industry are used for the evaluation. The nonrandom two-liquid segment activity coefficient (NRTL-SAC) and universal functional activity coefficient (UNIFAC) were selected to model the systems. The various thermodynamic relations existing in the open literature will be discussed and used to predict the solubility of active pharmaceutical ingredients and other small organic molecules in a single or a mixture of solvents. Equations of states, the activity coefficient, and predictive models will be discussed and used for this purpose. We shall also present some of our results on solvent screening using a single and a mixture of solvents.

Keywords: Solubility prediction, Pharmaceuticals, , NRTL-SAC Thermodynamic model, Activity coefficient, Solvent screening, Single solvent, Solvent mixture

1. Introduction

The study of solutions and their properties is one of the important branches of thermodynamics. It is important to know the behavior of a mixture of components for a change in temperature, pressure, and composition. Knowledge in this area of thermodynamics helps engineers

and scientists to better design, optimize, and operate the process units in the oil, gas, petrochemicals, pharmaceuticals, agricultural, and other chemical-related industries. Knowledge of the thermodynamics is essential to improve process performance and product quality. For example, the use of phase behavior calculations to understand and estimate the production rate of solids from a crystallization process in a pharmaceutical industry is of paramount importance in modeling, optimization, and control of product quality.

A solution is called ideal if its mixture property is a linear combination of the properties of each of its constituents at the given temperature and pressure [1]:

$$m_{solution} = \sum_{i=1}^{N} x_i m_i \tag{1}$$

where $m_{solution}$ and m_i represent the molar property of the mixture and pure component i, respectively, and x_i denotes the mole fraction of each constituent i in the solution. Not all solutions can be considered ideal. The interactions between the solute and solvent molecules in the solution renders a solution nonideal. The fundamental relationship which relates the thermodynamic properties is the Gibbs free energy [1]:

$$dG = VdP - SdT + \sum_{i=1}^{N} \left.\frac{\partial G}{\partial n_i}\right|_{P,T,n_j} dx_i \tag{2}$$

where G, V, P, S, and T are, respectively, Gibbs free energy, volume, pressure, entropy, and temperature of the solution. The $\frac{\partial G}{\partial n_i}\Big|_{P,T,n_j}$ represents the change in the Gibbs energy as a result of changes in the concentration of species i while the pressure, temperature, and the molar content of other species are kept constant. This is known as the chemical potential and in some textbooks is shown by μ_i. For a real solution, the chemical potential for each species is expressed by

$$\mu_i = \mu_i^{o} + RTlna_i \tag{3}$$

In which $\mu_i{}^o$ and a_i are the chemical potential of species i in its standard conditions and activity of the species i. The activity is defined as

$$a_i = \gamma_i x_i \tag{4}$$

where γ_i is the activity coefficient which is 1 for ideal solutions. Inserting Equation (4) into Equation (3), results in

http://dx.doi.org/10.5772/60982

$$\mu_i = \mu_i^o + RTln\gamma_i + RTlnx_i \tag{5}$$

The term $RTln\gamma_i$ accounts for the nonideality of the solution (it is also referred to as the partial molar energy). As it can be seen from Equation (5), the terms on the right side, except $Tln\gamma_i$, are calculated from the pure properties. However, in order to have an accurate prediction of the chemical potential of any species in the solution, $RTln\gamma_i$ should be known as well. In general, the activity coefficient is a function of temperature and composition and to a much less extent, the pressure. Because the activity coefficient is defined for a liquid solution, the pressure has very little effect on it. However, temperature and mole fractions of the species have significant effects on the activity coefficient of each species in a solution.

This chapter provides the reader with different and the most up-to-date thermodynamic models to estimate the activity coefficients of active pharmaceutical ingredients (API) in a solution.

1.1. Thermodynamics of the solutions containing dissolved solids

For a solid in equilibrium with itself in a solution, there is a thermodynamic relation [2]:

$$\hat{f}_i^s\left(T,P\right) = \hat{f}_i^L\left(T,P,x_i\right) \tag{6}$$

where $\hat{f}_i^s(T, P)$ denotes fugacity of component *i* in solid phase and $\hat{f}_i^L(T, P, x_i)$ represents the fugacity in the liquid phase. Equation (7) below correlates the fugacity of a pure component *i* in solid and liquid state (f_i^s and f_i^L, respectively).

$$f_i^s = x_i\gamma_i f_i^L \tag{7}$$

In order to find the ratio of the two fugacities, we need to consider all the thermodynamic processes that are involved from a solid to the liquid state. The three processes are shown in Figure 1. Path A in this figure shows the transformation from process conditions to the state where the solid starts melting. Path B shows the melting process at constant temperature and pressure. Path C indicates the change from melting to the process state. The sum of the three paths will give us the whole change from solid to liquid state of a pure component.

From the fundamental rules in thermodynamics, we have:

$$\Delta G_{s\to L} = RTln\frac{f_i^L}{f_i^S} \tag{8}$$

And the Gibbs energy change is related to change of enthalpy and entropy:

$$\Delta G_{s\to L} = \Delta H_{s\to L} - T\Delta S_{s\to L} \tag{9}$$

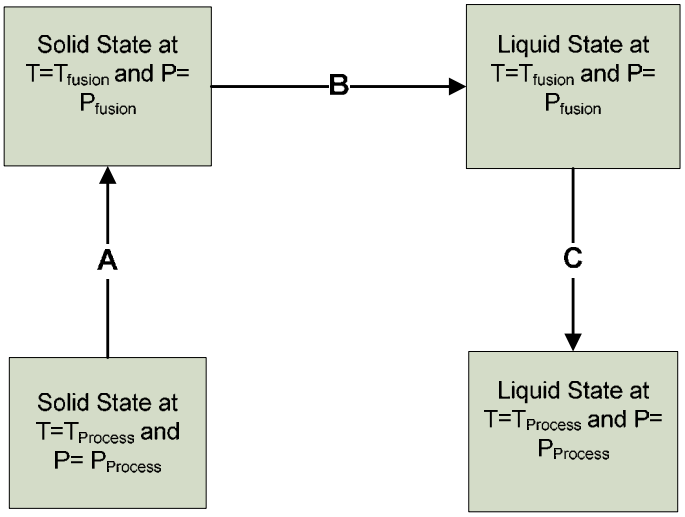

Figure 1. Schematic diagram for finding the fugacity change from solid to liquid state of a pure substance.

From Figure 1, the whole change in enthalpy is:

$$\begin{gathered} C_{p,solid}dT + \Delta H_{fusion} + \int_{T_f}^{T} C_{p,liquid}dT = \int_{T_f}^{T} \Delta C_p dT + \Delta H_{fusion} \\ \Delta H_{s\to L} = \Delta H_A + \Delta H_B + \Delta H_C = \int_{T}^{T_f} \end{gathered} \tag{10}$$

where $\Delta C_p = C_{p,liquid} - C_{p,solid}$. In the same manner, the entropy change from solid to liquid state can be found from

$$\Delta S_{s\to L} = \int_{T_f}^{T} \frac{\Delta C_p}{T} dT + \Delta S_{fusion} \tag{11}$$

Also, $\Delta S_{fusion} = \frac{\Delta H_{fusion}}{T_{fusion}}$. If we substitute Equations (10) and (11) into Equation (9), then:

$$RTln\frac{f_i^L}{f_i^S} = \int_{T_f}^{T} \Delta C_p dT + \Delta H_{fusion} + \int_{T_f}^{T} \frac{\Delta C_p}{T} dT + \frac{\Delta H_{fusion}}{T_{fusion}} \tag{12}$$

http://dx.doi.org/10.5772/60982

If we neglect the terms including change in the heat capacity (because of the large value of heat of fusion compared to the heat capacities), then we will get the following important equation:

$$ln x_s = \frac{-\Delta H_{fus}}{R}\left(\frac{1}{T_m} - \frac{1}{T}\right) - ln\gamma_s \tag{13}$$

where x_s is the equilibrium mole fraction of the solute (dissolved solid) in a solution at temperature T. Equation (13) is the starting point in every solid-liquid equilibrium calculation procedure that relates the three important variables x_s, T, and γ_s. However, we already know that γ_s is a function of solution temperature and the mole fraction of the species. Therefore, one can fix all the thermodynamic properties of a solution if the mole fraction of the species and the solution temperature are known. We recall that although the pressure has to be fixed as well, but due to its minimal effect on the activity coefficient and other properties of the system, we don't need to have it for calculations.

1.2. Classification of the thermodynamic models for solutions

In order to predict the solubility of solids in pure and mixed solvents, the use of noncorrelative (predictive) thermodynamic models is of high importance. Since several years ago, there have been many thermodynamic models introduced to help scientists and engineers in the prediction of phase behaviors of various liquid-liquid and vapor-liquid systems, such as the Margules equation [3], the Wilson equation [4], the Van Laar equation [5], the nonrandom two-liquid (NRTL) equation [6], and the UNIQUAC equation [7]. As these models are applicable in the estimation of activity coefficients, they can also be utilized to predict solid-liquid equilibrium systems. In general, the models which govern the activity coefficients of the species in the solutions are grouped into two categories:

1. The models which need experimental data at various temperatures, pressures, and compositions of the mixture, such as UNIQUAC equation.

2. The models which only need some fundamental properties of the chemical molecule and a very few experimental data to predict the phase behavior of the solids within various solvents, such as the UNIFAC [8] and NRTL-SAC models [9].

As it is apparent from the above two categories, the first one is not useful for the prediction of the phase behavior of mixtures, specifically for mixtures of more than two species. Because of the time-consuming and expensive nature of binary interaction parameter evaluation of various chemical components, the first group of thermodynamic models (above) is not practical. As an example, the activity coefficient from Wilson's equation of state is found from [4]:

$$ln\gamma_i = -ln\left[\sum_{j=1}^{n} x_j \Lambda_{ij}\right] + 1 - \sum_{z=1}^{n} \frac{x_z \Lambda_{ki}}{\sum_{k=1}^{n} x_k \Lambda_{zk}} \tag{14}$$

The binary interaction parameter is:

$$\Lambda_{ij} = \frac{V_j^L}{V_i^L} exp\left[\frac{-\lambda_{ij} - \lambda_{ii}}{RT}\right] \tag{15}$$

where *i* and *j* refer to the compounds present in the solution. From Equation (14), it can be seen that for obtaining the activity coefficient of a component 1 in a pure solvent 2, we need four interaction parameters (Λ_{12}, Λ_{21}, $\Lambda_{11} \wedge \Lambda_{22}$, which are temperature dependent. In addition, from Equation (15), it is evident that for calculating the value of the binary interaction parameters, additional experimental data, such as molar volume is needed.

From the predictive category (second group), the universal functional activity coefficient (UNIFAC) model is a well-known example. The main application of the UNIFAC model is in systems showing nonelectrolytic and nonideal behaviors. Fredenslund et al. [8] developed the UNIFAC model. The NRTL segment activity coefficient (NRTL-SAC) model was first introduced in 2004 by Chen et al. [9]. This model was proposed in order to compensate for the weakness of the UNIFAC in predicting the solubility of complex chemical molecules with functional groups that had not been studied for the UNIFAC parameters. Also, in some cases, the UNIFAC group addition rule becomes invalid [2]. One of the main advantages of the NRTL-SAC model in comparison to the other predictive methods is its ability to predict organic electrolyte systems. The UNIFAC method identifies the molecule in terms of its functional groups, while the NRTL-SAC model divides the whole surface of the molecule to four segments. These segments' values are found from their interactions with other molecules in a solution. Based on Chen et al., three purely hydrophilic, hydrophobic, and polar segments have been selected as water, hexane, and acetonitrile, respectively.

2. Ideal solutions

An ideal solution is simply defined as a mixture of chemical components with its thermodynamic property related to the linear sum of each pure species thermodynamic property (Equation (1)). A common example is a solution which obeys Raoult's law. This law states that the total pressure of a system is a linear combination of the component's vapors pressure at the system's temperature, provided that the total pressure is less than 5 atm. In order to derive Raoult's law, we start from Equation (5) and assume that the liquid solution is ideal:

$$\mu_i = \mu_i^o + RTlnx_i \tag{16}$$

If Equation (16) is written for component *i* in both the liquid and vapor phases, we have:

http://dx.doi.org/10.5772/60982

$$\mu_i^L = \mu_i^{o,L} + RTlnx_i \tag{17}$$

$$\mu_i^V = \mu_i^{o,V} + RTln\frac{\hat{f}_i}{P} \tag{18}$$

where $\hat{f}_i$ and P are the fugacity of species i in the vapor mixture and the total pressure, respectively. From the thermodynamic equilibrium criteria, if the two phases of liquid and vapor are in a state of equilibrium, then:

$$\mu_i^L = \mu_i^V \tag{19}$$

After substitution of Equations (17) and (18) into Equation (19), we get:

$$\mu_i^{o,L} + RTlnx_i = \mu_i^{o,V} + RTln\frac{\hat{f}_i}{P} \tag{20}$$

If Equation (19) is written for the case of pure component i in liquid phase at equilibrium with its vapor phase, then:

$$\mu_i^{o,L} = \mu_i^{o,V} + RTln\frac{f_i}{P} \tag{21}$$

By comparing Equations (20) and (21), it yields:

$$RTlnx_i = RTln\frac{\hat{f}_i}{f_i} \tag{22}$$

Equation (22) is simplified to:

$$\hat{f}_i = f_i x_i \tag{23}$$

Now, if the liquid and vapor mixtures are assumed to be ideal, then the fugacity values can be substituted by their corresponding pressure and thus:

$$P_i = P_i^{sat} x_i \tag{24}$$

P_i is called the partial pressure of species i in the vapor mixture and P_i^{sat} denotes the pure vapor pressure of the component i at the solution temperature.

2.1. Ideal solution mixtures: VLE phase behavior

In order to calculate the properties of a system of components that obey Raoult's law, Equation (24) is written for each of the species present in the solution,

$$\sum_{i=1}^{N} y_i P_{tot} = \sum_{i=1}^{N} P_i^{sat} x_i \tag{25}$$

and by simplification:

$$P_{tot} = \sum_{i=1}^{N} P_i^{sat} x_i \tag{26}$$

by which the solution equilibrium temperature can be found. As we know, the vapor pressure of each species is temperature-dependent and by having the mole fraction of the components in the liquid phase, one can find the temperature which corresponds to the total pressure of the system. This problem can be solved quickly using a spreadsheet and by changing the solution temperature until Equation (26) is satisfied. Sometimes, the temperature of the solution is known and the total pressure needs to be calculated, which is a straightforward calculation from Equation (26). This way, there is no need to use the nonlinear solvers to find the temperature.

There are two examples of the binary systems which exhibit negative deviation from Raoult's law.

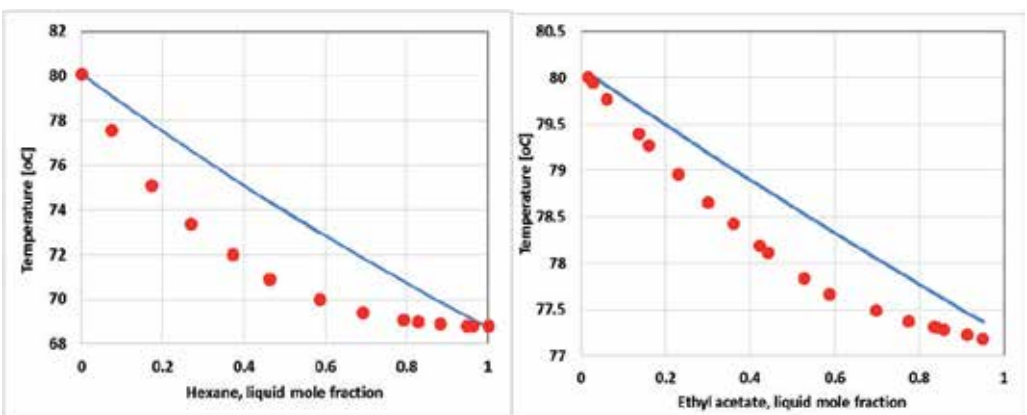

Figure 2. T-x-y diagram for the binary systems of hexane-benzene (left) and ethyl acetate-benzene (right) at a pressure of 101.33 kPa.

http://dx.doi.org/10.5772/60982

In order to better demonstrate whether a system follows Raoult's law, a diagram of the phase equilibrium called T-x-y should be plotted. This plot (Figure 2) shows the equilibrium temperatures at which either a liquid solution will start bubbling (bubble curve) or a vapor mixture starts condensing (dew curve). The two systems with their experimental data and the calculation curve of the ideal solution is shown in Figure 2. In Figure 2, the system of hexane-benzene at the pressure of 101.33 kPa [10] and the system of ethylacetate-benzene [11] show negative deviations from Raoult's law.

If one is interested in finding the bubble and dew curves for an ideal solution at low to moderate pressures, then:

- **Bubble point**

From Equation (26), by knowing the mole fraction of each component in the liquid phase, two cases are possible:

1. **Pressure known.** If the total pressure of the system is given, then Equation (26) should be solved for the temperature (if the vapor pressure follows Antoine's law):

$$P_i^{sat} = A - \frac{B}{T + C}, P_{tot} = \sum_{i=1}^{N} P_i^{sat}(T) x_i \tag{27}$$

2. **Temperature known.** For this case, the problem is straightforward. The vapor pressure of each species can be found readily from Antoine's equation and then inserted into Equation (27) to find the total pressure.

- **Dew point**

1. **Pressure known.** If Raoult's law is isolated for the mole fraction of the species in the vapor phase, then:

$$y_i = \frac{P_i^{sat} x_i}{P_{tot}} \tag{28}$$

by taking the sum for all the components in the liquid phase and knowing that the sum of the mole fractions in a phase is 1, we get:

$$1 = \sum_{i=1}^{N} x_i = \frac{y_i P_{tot}}{P_i^{sat}(T)} \tag{29}$$

If the total pressure is known, then Equation (29) should be solved for the temperature which determines the vapor pressure of each of the species in the mixture.

2. **Temperature known.** In this case, Equation (29) is isolated for the P_{tot} to find the total pressure of the system:

$$P_{tot} = \frac{1}{\sum_{i=}^{N} \frac{P_i^{sat}(T)}{y_i}} \tag{30}$$

which has a straightforward solution.

2.2. Ideal solution mixtures: SLE phase behavior

For a solid-liquid equilibrium behavior, Equation (13) for the ideal solution (in which γ_s=1) is simplified to:

$$lnx_s = \frac{-\Delta H_{fus}}{R}\left(\frac{1}{T_m} - \frac{1}{T}\right) \tag{31}$$

From Equation (31), one can find the mole fraction (or what is called *solubility* in the case of SLE) by inserting the appropriate values for the enthalpy of fusion, ΔH_{fus}, melting temperature, T_m, and the solution temperature, T. It can be found from this equation that for any solvent, the solubility of a solid will be the same regardless of the nature of the solvent. For very few cases, this assumption might be correct; however, for most of the practical and common applications, the above formula does not work well. Therefore, a term accounting for the nonideality of the system should be added to Equation (31).

3. Nonideal solutions

Almost all real VLE or SLE systems show nonideality. Equations that predict the behavior of different phase equilibria are divided into two:

- Equations of state (EOS) that are useful to predict the nonideal behavior of the vapor phase
- Equations which are applied to the liquid phase to predict the nonideal behavior of the liquid solutions

We focus on the thermodynamic models that deal with the liquid mixtures in this chapter. From the two categories of activity coefficient models, the correlative one is not very useful for solubility prediction and solvent screening purposes. The main reason for this is the lack of experimental data for the binary interaction parameters of the solute-solvent, solute-antisolvent, and solvent-antisolvent systems. As an example, the activity coefficient from

http://dx.doi.org/10.5772/60982

Wilson's equation of state is found from Equations (14) and (15). It can be seen that for obtaining the activity coefficient of a component 1 in a pure solvent 2, we need four interaction parameters (Λ_{12}, Λ_{21}, $\Lambda_{11} \wedge \Lambda_{22}$, which are temperature dependent. It is evident that for calculating the value of the binary interaction parameters, additional experimental data, such as molar volume is needed. Other models which belong to the first category have the same limitations as Wilson's method. The Wilson model was used in the prediction of various hydrocarbons in water in pure form and mixed with other solvents by Matsuda et al. [11]. In order to estimate the pure properties of the species, the Tassios method [12] with DECHEMA VLE handbook [13] were used. Matsuda et al. also took some assumptions in the estimation of binary interactions (because of the lack of data) that resulted in some deviations from the experimental data.

From the predictive category, we bring some examples of the application of the UNIFAC model. In one study, this model has been used to predict the solubility of naphtalene, anthracene, and phenanthrene in various solvents and their mixtures [8]. They showed the applicability of the UNIFAC model in prediction of the phase behavior of solutes in solvents. There have been efforts to make the UNIFAC model more robust and powerful in the prediction of phase behaviors [14]. In one study, the solubility of buspirone-hydrochloride in isopropyl alcohol was measured and evaluated by the modified UNIFAC model [15]. It was concluded that for highly soluble pharmaceutics, the modified form of the UNIFAC model was not suitable. In another study, the solubility of some chemical species in water and some organic solvents was predicted by the UNIFAC model [16]. For some unknown functional groups, they used other known groups which had chemical structures that were similar to unknown ones. In conclusion, it was stated that the UNIFAC model is not a proper model for use in crystallization and related processes. The UNIFAC model also has been utilized to predict the solubility of some aromatic components as well as long-chain hydrocarbons [17]. The results showed that the predictions for the linear hydrocarbons are not as good as the ones for the aromatics.

Chen et al. developed the NRTL-SAC model in 2004 [9]. One of the main reasons for developing this model was to enhance the predicting capability of the solubility of complex chemical molecules with functional groups that were not included in the UNIFAC model. Also, in some cases, the UNIFAC group addition rule becomes invalid [2]. One of the main advantages of the NRTL-SAC model in comparison to the other predictive methods is its ability to predict organic electrolyte systems [18]. The UNIFAC method identifies the molecule in terms of its functional groups, while the NRTL-SAC model divides the whole surface of the molecule to four segments. The hydrophobic segment (X) denotes parts of the molecule surface which don't participate in forming a hydrogen bond, such as hexane. The polar segments (Y- and Y+) do not belong to either hydrophobic or hydrophilic segment. The polar attractive segment (Y-) shows attractive interaction with hydrophilic segment, while the polar repulsive segment (Y +) has repulsive characteristic with hydrophilic segment. Hydrophilic segment (Z) contributes to the part of the molecule which tends to form a hydrogen bond, such as water.

Based on the interaction forces between the molecule surfaces, the four segments of the NRLT-SAC model have been identified. The three reference solvents mentioned before were used to

determine the rest of chemical components' segment numbers [9]. As an example, the VLE and LLE data containing the component of interest in binary mixtures with hexane, water, and acetonitrile were collected and then used in order to find the segment numbers of the component. Sheikholeslamzadeh et al. have investigated various industrial case studies to show the applicability of the NRTL-SAC in processes consisting of solvent mixtures [19]. The segment numbers for each of the nonreference solvents are found from the nonlinear optimization methods which minimize the deviation from the model output and the experimental data for the whole range of data (VLE and LLE). Most of the solvents have some segment numbers that are high in value compared to other segment numbers. The reason is that it is less probable that a chemical component collects all four segments at the same level of value. Currently, more than a hundred solvent segment numbers have been evaluated and optimized which can be found in some commercial simulation packages. The two important predictive equations of UNIFAC and NRTL-SAC, as representatives of the activity coefficient models, are presented here.

3.1. UNIFAC model

In general, the activity coefficient models of the predictive category split the activity coefficients into two segments:

- A part that includes the contribution of the chemical structure and the size of a compound (combinatorial part)
- The second part that includes the contribution of the functional sizes and binary interaction between pairs of the functional groups (residual part)

With the above definition, the total activity of a component in the solution is the sum of the two parts:

$$ln\gamma_i = ln\gamma_i^C + ln\gamma_i^R \tag{32}$$

In which γ_i is the activity coefficient of component i in the solution, $\gamma_i{}^C$ is the combinatorial part and $\gamma_i{}^R$ is the residual part. Up to this point, all of the group contribution and activity coefficient methods (i.e. NRTL-SAC) have been the same, but the methods in which the activities have been calculated are different. In the UNIFAC model, the combinatorial part for component i is found from the following equation [8]:

$$ln\gamma_i^c = ln\left[\frac{\phi_i}{x_i}\right] + \frac{z}{2}q_i ln\left[\frac{\theta_i}{\phi_i}\right] + L_i - \frac{\phi_i}{x_i}\sum_{j=1}^{n} x_j L_j \tag{33}$$

where

$$L_i = \frac{z}{2}\left(r_i - q_i\right) - \left(r_i - 1\right) \tag{34}$$

http://dx.doi.org/10.5772/60982

z is the coordination number and is taken to be 10. In Equation (33), ϕ_i is the segment fraction and θ_i is the area fraction of component i and is related to the mole fraction of species i in the mixture:

$$\phi_i = \frac{r_i x_i}{\sum_{j=1}^{n} r_j x_j} \tag{35}$$

$$\theta_i = \frac{q_i x_i}{\sum_{j=1}^{n} q_j x_j} \tag{36}$$

q_i and r_i are the pure component surface area and volumes (van der Waals), respectively. These parameters are not temperature-dependent and are only functions of the chemical structure of a functional group. In the UNIFAC model, for every functional group, there is a unique value for surface area and volume that can be found in common texts and handbooks [1]. The first step in modeling the UNIFAC for a specific binary or ternary system is to break down the chemical structure of a molecule into the basic functional groups. As it is suggested in thermodynamic textbooks [19], the optimum way of breaking it down is the one which results in the minimum number of subgroups, and with each subgroup having the maximum replicates. The q_i and r_i can be found from Equations (37) and (38):

$$r_i = \sum_k v_k{}^i R_k \tag{37}$$

$$q_i = \sum_k v_k{}^i Q_k \tag{38}$$

$v_k{}^i$ is the number of subgroup k in component i. The residual part of the UNIFAC is found from the equation below:

$$ln\gamma_i{}^R = \sum_{k=1}^{n_k} v_k{}^i \left[ln\Gamma_k - ln\Gamma_k{}^i \right] \tag{39}$$

In Equation (39), Γ_k is the residual activity coefficient of subgroup k in the mixture and $\Gamma_k{}^i$ is that value in a pure solution of the component i. This term is added so when the mole fraction approaches unity, the term $ln\gamma_i{}^R$ tends to zero $(\gamma_i{}^R \to 1)$. The residual activity coefficient of subgroup k in a solution is given by

$$ln\Gamma_k = Q_k\left[1 - ln\sum_m \theta_m \psi_{mk} - \sum_m\left[\frac{\theta_m \psi_{mk}}{\sum_n \theta_n \psi_{nm}}\right]\right] \tag{40}$$

Equation (40) is also applicable to the case of $\Gamma_k{}^i$, in which the parameters of the right-hand side of the equation are written based on the pure component *i*. θ_m is the area fraction of the functional group *m* in the mixture:

$$\theta_m = \frac{Q_m X_m}{\sum_n Q_n X_n} \tag{41}$$

X_m is the mole fraction of subgroup *m* in the mixture. ψ_{nm} is the group interaction parameter between groups *n* and *m* and is dependent on the temperature:

$$\psi_{nm} = exp\left[\frac{-u_{nm} - u_{mm}}{RT}\right] = exp\left[\frac{-a_{nm}}{T}\right] \tag{42}$$

The group interaction parameter a_{nm} is found from the large sets of VLE and LLE data in the literature, which are tabulated for many subgroups. It is worth noting that $a_{nm} \neq a_{mn}$. There are some modifications to the original UNIFAC equation in order to make the model robust for some complex systems. In the UNIFAC-DM method, the modification is made on the combinatorial part:

$$ln\gamma_i^{\,c} = ln\left[\frac{\phi_i'}{x_i}\right] + \frac{z}{2} q_i ln\left[\frac{\theta_i}{\phi_i}\right] + L_i - \frac{\phi_i}{x_i}\sum_{j=1}^{n} x_j L_j \tag{43}$$

In which the term $\frac{\phi_i'}{x_i}$ is defined as

$$\frac{\phi_i'}{x_i} = \frac{r^{3/4}}{\sum_j x_j r_j^{3/4}} \tag{44}$$

The algorithm for finding the solute solubility in the mixture of ternary solution (solvent/cosolvent/solute) is shown in Figure 3. The algorithm starts with known values, such as the physical properties of the solute. After making an initial guess for the solubility, the program obtains the activity coefficients and the new solubility is found and is compared with the old

http://dx.doi.org/10.5772/60982

value and the calculations are repeated to converge to a unique value for solubility. This procedure is done for all the experimental data points.

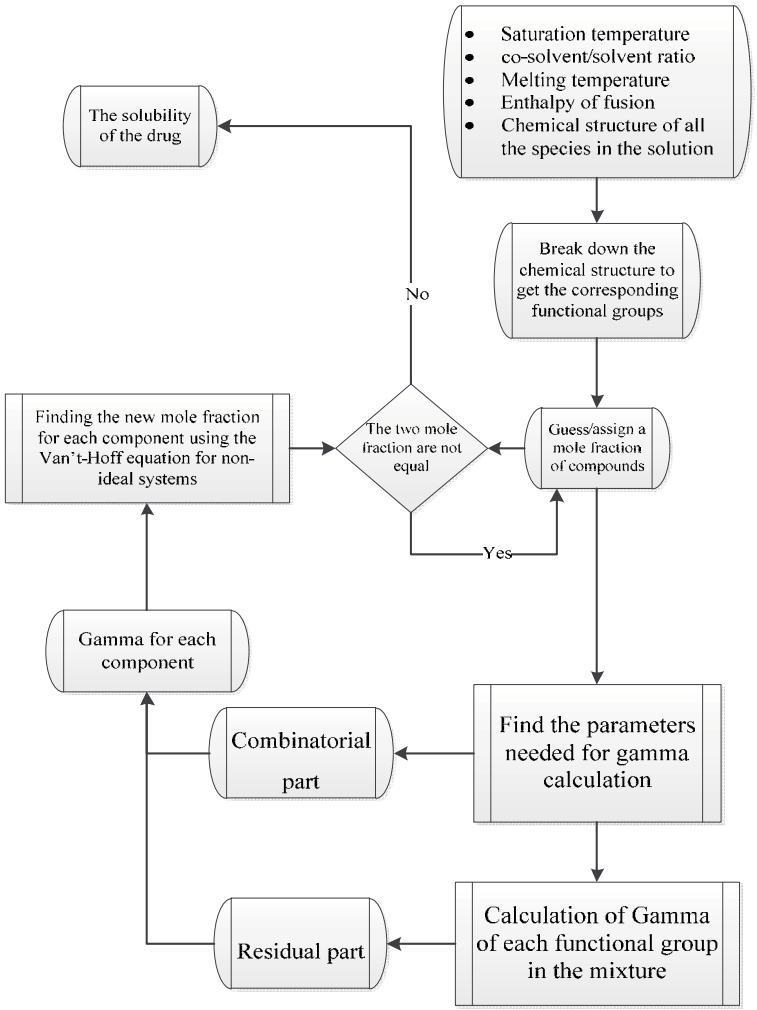

Figure 3. The algorithm of converging to the solubility of a ternary system using the UNIFAC model.

3.2. Nonrandom two-liquid segment activity coefficient (NRTL-SAC)

According to Chen et al. [9], the NRTL-SAC model is based on the derivation of the original NRTL model for polymers. From Equation (32), the activity coefficient is made up of two terms, combinatorial and residual. Like the UNIFAC model, the activity coefficients must be generated in order to obtain solubility. In the NRTL-SAC model, the combinatorial part is calculated by Equation (45):

$$ln\gamma_i^C = ln\frac{\varnothing_i}{x_i} + 1 - r_i\sum_j \frac{\varnothing_j}{x_j} \tag{45}$$

With the definitions:

$$r_i = \sum_j r_{j,i} \tag{46}$$

$$\varnothing_i = \frac{r_i x_i}{\sum_j r_j x_j} \tag{47}$$

where x_i is the mole fraction of component *i*, $r_{m,i}$ is the number of segment *m*, r_i is the total segment number in component *i*, and $\varnothing_i$ is the segment mole fraction in the mixture.

The residual term is defined as

$$ln\gamma_i^R = ln\gamma_i^{lc} = \sum_m r_{m,i}\left(ln\Gamma_m^{lc} - ln\Gamma_m^{lc,i}\right) \tag{48}$$

In Equation (48), there are two terms, $ln\Gamma_m^{lc}$ and $ln\Gamma_m^{lc,i}$, which are the activity coefficients of segment *m* in solution and component *i*, respectively.

The two mentioned terms are found using Equations (49) and (50):

$$ln\Gamma_m^{lc} = \frac{\sum_j x_j G_{j,m}\tau_{j,m}}{\sum_k x_k G_{k,m}} + \sum_{m'} \frac{x_{m'} G_{m,m'}}{\sum_k x_k G_{k,m'}}\left(\tau_{m,m'} - \frac{\sum_j x_j G_{j,m'}\tau_{j,m'}}{\sum_k x_k G_{k,m'}}\right) \tag{49}$$

$$ln\Gamma_m^{lc,l} = \frac{\sum_j x_{j,l} G_{j,m}\tau_{j,m}}{\sum_k x_{k,l} G_{k,m}} + \sum_{m'} \frac{x_{m',l} G_{m,m'}}{\sum_k x_{k,l} G_{k,m'}}\left(\tau_{m,m'} - \frac{\sum_j x_{j,l} G_{j,m'}\tau_{j,m'}}{\sum_k x_{k,l} G_{k,m'}}\right) \tag{50}$$

In the two above equations, *l* is referred to as the component and *j, k, m,* and m' are referred to the segments in each component. $x_{j,l}$ is the segment-based mole fraction of segment species *j* in component *l* only. The mole fractions of segments in the whole solution and in components are defined as below:

$$x_j = \frac{\sum_l x_l r_{j,l}}{\sum_z \sum_i x_z r_{j,z}} \tag{51}$$

http://dx.doi.org/10.5772/60982

$$x_{j,l} = \frac{r_{j,l}}{\sum_i r_{j,l}} \tag{52}$$

$G_{i,j}$ and $\tau_{i,j}$ are the local binary values which can be related to each other based on NRTL nonrandom parameter α_{ij}, and are shown by their values in Table 1. $G_{i,j}$ and $\tau_{i,j}$ have the following relation:

$$G_{i,j} = e^{-\alpha_{i,j}\tau_{i,j}} \tag{53}$$

Therefore, from fixed values of $\tau_{i,j}$ and $\alpha_{i,j}$ one can find $G_{i,j}$. The segment numbers for the common solvents can be found from the literature [20]. After putting the values of segments for solvents and initial guess values for the solute segments, the written code for NRTL-SAC starts solving for the mole fractions at saturation for all of the species in the solution (see Figure 3).

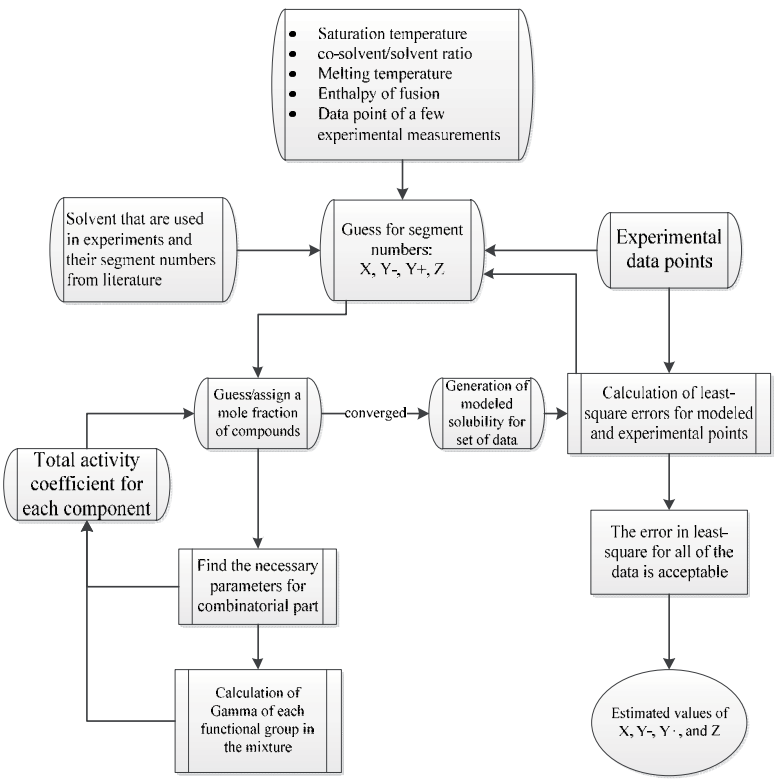

Figure 4. Algorithm flowchart for parameter estimation using NRTL-SAC model.

It is worth noting that the main difference in Figures 3 and 4 is the use of parameter estimation method for the calculation of the NRTL-SAC parameters, while for the UNIFAC model, the calculation is straightforward. Once the parameters (here, the segment numbers) are found, then they could be used for validation against other experimental data.

4. Application of solution thermodynamics in industry

One of the main applications of the thermodynamic models is in the chemical industries which use solvent (or their mixtures) [19-22]. Two cases of the vapor-liquid equilibrium of common industrial solvent systems are discussed here.

4.1. VLE study of two binary azeotropic systems

There are some solvents within the chemical and pharmaceutical industries which are of importance to study, such as ethanol, isopropyl alcohol, and water in addition to some aromatic components, such as benzene and toluene. These solvents are sometimes used as additives to other valuable chemicals to maintain some performances or enhance their physical or chemical properties. The systems that form azeotropes make the distillation process calculations complex. As a result, having an accurate knowledge about the phase behaviors of such systems are important (such as toluene-ethanol or toluene-isopropyl alcohol). The four case studies of the VLE data for the mentioned azeotropic binary systems were used in a study by Chen et al. [24]. The operating pressures were at four distinct levels for the calculations. They have used three equations of state (from the correlative group) to fit the experimental data with the model outputs and found the necessary binary interaction parameters. Based on their work, the estimated parameters were found; however, they were not used to further validate the models for other operating conditions. In a study by Sheikholeslamzadeh et al., they used the NRTL-SAC and UNIFAC models to predict these azeotropic systems [20]. Table 1 (from their work) shows average relative deviation for the compositions and temperature for the two systems. It is seen from the table that the deviations from the experimental data in the vapor phase mole fractions are almost one-third relative to the NRTL-SAC model. On the other hand, the relative deviation for the saturation temperature is higher using the NRTL-SAC model. Another fact from this finding is that the deviations for both binary systems when the NRTL-SAC model is used are nearly the same. This is not the case when utilizing the UNIFAC model to predict the systems' behaviors. The final conclusion from Table 1 would be the better predictive capability of the UNIFAC model for light alcohols than the heavier ones.

Thermodynamic model	Pressure, (kPa)	%ARD (equilibrium temperature and vapor-phase mole fractions)					
		System 1			System 2		
		Temperature, (K)	Ethanol	Toluene	Temperature, (K)	2-Propanol	Toluene
NRTL-SAC	101.3	0.46	5.14	9.81	0.41	4.91	7.19

http://dx.doi.org/10.5772/60982

Thermodynamic model	Pressure, (kPa)	%ARD (equilibrium temperature and vapor-phase mole fractions)					
		System 1			System 2		
		Temperature, (K)	Ethanol	Toluene	Temperature, (K)	2-Propanol	Toluene
UNIFAC	201.3	0.64	7.63	9.72	0.66	7.76	9.77
	101.3	0.10	1.08	3.15	0.22	2.04	3.87
	201.3	0.27	2.03	3.47	0.36	3.43	5.40

Table 1. The results of the prediction using the NRTL-SAC and UNIFAC models with the experimental values of Chen et al. [20,24]

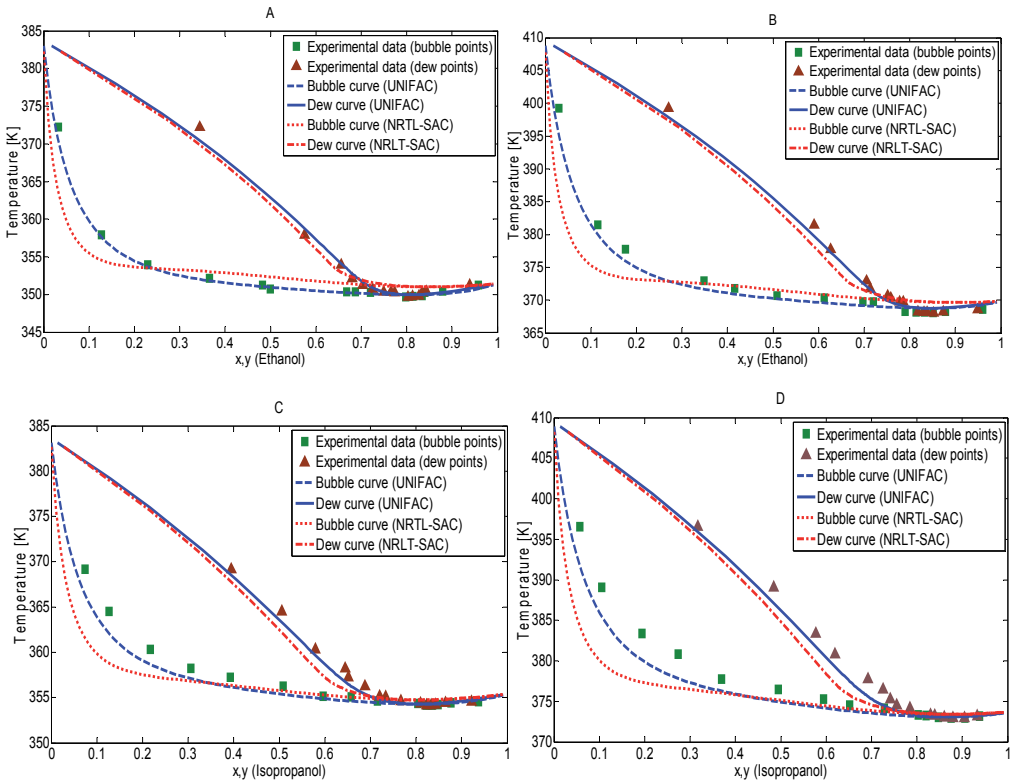

Figure 5. The results for the systems (A) ethanol-toluene at 101.3 kPa, (B) ethanol-toluene at 201.3 kPa, (c) isopropyl alcohol-toluene at 101.3 kPa, and (D) isopropyl alcohol-toluene at 201.3 kPa [20].

4.2. VLE study of a ternary system

With the increasing prices of fossil fuels, the demand to obtain alternatives has received much attention. One of the candidates for this purpose is ethanol, as it decreases the air pollution when blended to the conventional fossil fuels and thereby, increasing the performance of burning fuels within the vehicle engines. It would also be cost-effective when adding other

low-price additives to the fuel. The higher the purity of the alcohol being used as additive, the better the performance of the fuel. It was found that the addition of glycols to the mixture containing alcohols and water can improve the separation processes and utilize less energy to perform the process [25,26]. The experimental data containing the ternary system of ethylene glycol-water-ethanol and the performance of separation by varying the glycol concentration were performed by Kamihama et al. [27].

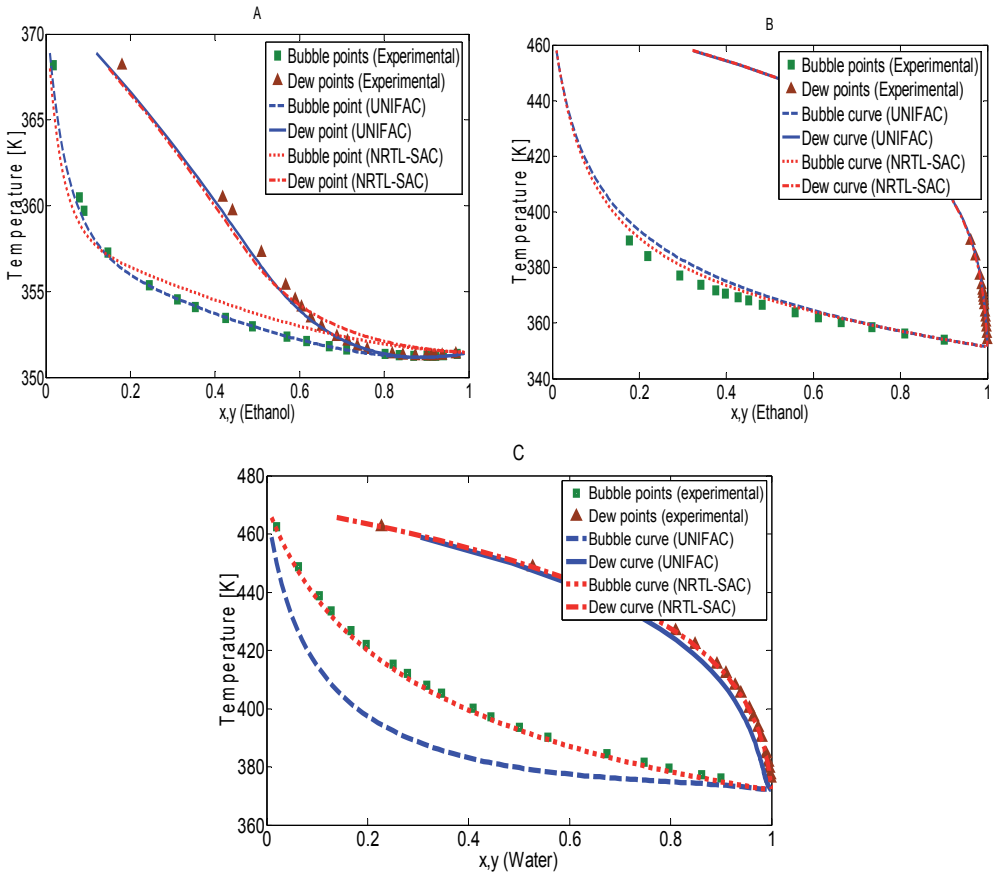

Figure 6. VLE diagrams of the systems (A) ethanol-water, (B) ethanol-ethylene glycol, and (C) water-ethylene glycol at 101.3 kPa [20].

They performed binary system experiments for each of the pairs in the ternary system. It was found that glycol can move the azeotrope point and therefore, enhance the separation process. Sheikholeslamzadeh et al. have used both the NRTL-SAC and UNIFAC models to perform phase calculations and assess the capacity of the mentioned models in the prediction of binary and ternary systems containing glycol and alcohols [20].

From Figure 6, the UNIFAC model has the capability of predicting the system of ethanol-water, perfectly. However, this is not the case for the systems using ethylene glycol-water and

http://dx.doi.org/10.5772/60982

ethylene glycol-ethanol. On the other side, the NRTL-SAC model gave satisfactory results for all three binary system, specifically, the systems that contain ethylene glycol. Also from Figures 6B and 6C, it can be seen that for nonazeotropic systems, the NRTL-SAC model could best show capability in prediction comparable to the UNIFAC model. The UNIFAC model could best locate the azeotrope point and the VLE behavior of the ethanol-water system.

For the ternary system of solvents consisting of ethanol, water, and ethylene glycol, Kamihama et al. [27] conducted the vapor-liquid measurements at a pressure 101.3 kPa. In order to use the correlative models (such as Wilson) for the ternary system, the binary interaction parameters should be known for each pair of components in the mixture at that temperature and pressure. They used this method to find the ternary behavior of the system. Sheikholeslamzadeh et al. used the NRTL-SAC model with the four conceptual segments of each solvent, which were already accessible in the literature [9, 18]. The results showed the high performance of the NRTL-SAC model in the prediction of this ternary system. The bubble point temperature as well as the vapor and liquid compositions could be estimated fairly well with the NRTL-SAC model. The predicted results are shown in Figure 7, giving the vapor phase mole fractions of the three species as well as the mixture temperature. It is apparent that the deviation from the experimental data for the case of water and ethanol in this ternary mixture is almost zero. The UNIFAC model could not match the experimental data as well as the NRTL-SAC model. For the ethylene glycol, as the experimental concentrations of the vapor phase are very small, with the inclusion of errors of the experimentation, the NRTL-SAC model could also give satisfactory results. Finally, for the bubble point temperature, Figure 7 illustrates the perfect predictions of the NRTL-SAC model compared with the UNIFAC model. The average relative deviation for the whole set of experimental data on the saturated temperature from the NRTL-SAC model is one-third that from the UNIFAC model.

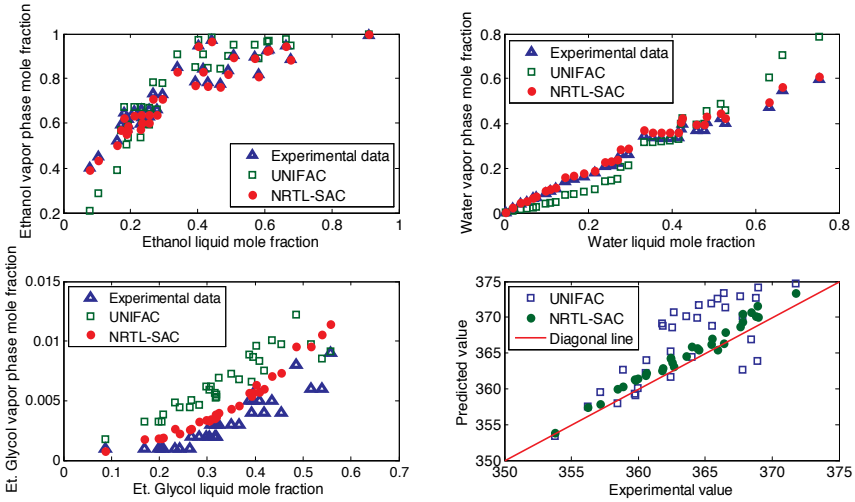

Figure 7. Vapor phase mole fractions of ethylene glycol (bottom-left), ethanol (top-left), water (top-right), and the bubble temperature (bottom-right) prediction from the NRTL-SAC model [20].

5. Conclusion

There are several equations of state that describe the phase behavior of chemical components of a system at various temperatures, pressures, and compositions. From these models, the first group which needs various experimental data to predict the system behavior at other conditions is not very attractive. Instead, the second group (predictive models) is based on the activity coefficients that are found from the molecular structures with a few experimental data. In this chapter, the capacity of handling two binary and ternary systems of solvents using those predictive models was assessed. The NRTL-SAC and UNIFAC models were chosen for the modeling of those systems. The NRTL-SAC model showed relative advantage over the UNIFAC model in almost all cases, except for the systems containing light alcohols with water. The preference of using NRTL-SAC is due to its simplicity compared to the UNIFAC. If the four segment parameters of a specific component are known, then they can be set as a unique value for that component irrespective of the mixture conditions. This way, various operating conditions can be defined and the phase behavior of the components can be predicted accurately and rapidly. One of the main advantages of the NRTL-SAC model is that it can be written in various computer programming languages to be used in process simulation analysis. One good example of such work can be found in Sheikholeslamzadeh et al. [19,21]. They used the NRTL-SAC model to find the solid-liquid equilibrium for three pharmaceuticals. Then, the parameters they optimized for the given pharmaceuticals were used further to perform a solvent screening method. This method is really costly and time-consuming in industry. The authors [19,20] have developed an algorithm to find the best combination of solvents, temperatures, and pressures for the best yield of pharmaceutical production using the NRTL-SAC model. In conclusion, the NRTL-SAC model and other similar ones open a new window for the engineers and scientists to have a wider, more accurate, and rapid predictions of the solubility of active pharmaceuticals in mixtures of solvents.

Author details

Ehsan Sheikholeslamzadeh and Sohrab Rohani*

*Address all correspondence to: Srohani@uwo.ca

Department of Chemical and Biochemical Engineering, Western University, Canada

References

[1] J. M. Smith, Hendrick C Van Ness, Michael Abbott. Introduction to Chemical Engineering Thermodynamics, McGraw-Hill Science, Sixth Edition, 2004.

[2] J. M. Prausnitz, R. N. Lichtenthaler, E. G. Azevedo. Molecular Thermodynamics of Fluid-Phase Equilibria, Prentice-Hall International Series, Third Edition, 1999.

[3] M. Margules, Uber die Zusammensetzung der gesattigten Dampfe von Mischungen, Sitzungsber Akad. Wiss. Wien. Math. Naturw. Klasse, II, vol. 104, pp. 1243-12778, 1985.

[4] G. M. Wilson. Vapor-liquid equilibirium. XI. A new expression for the excess free energy of mixing. J. Am. Chem. Soc., vol. 86, pp.127-130, January 1964.

[5] J. J. van Laar. Sechs Vortragen uber das thermodynamische potential (Six Lectures on the Thermodynamic Potential). Braunschweig, Fried. Vieweg & Sohn, 1906.

[6] H. Renon and J. M. Prausnitz. Local compositions in thermodynamic excess functions for liquid mixtures. AIChE J., vol. 14, pp. 135-144, January 1968.

[7] D. S. Abrams and J. M. Prausnitz. Statistical thermodynamics of liquid mixtures: A new expression for the excess Gibbs energy of partly or completely miscible systems. AIChE J., vol. 21, pp. 116-128, January 1975.

[8] A. Fredenslund, R. L. Jones, J. M. Prausnitz. Group contribution estimation of activity coefficients in nonideal liquid mixtures. AIChE J., 1975, 21 (6), 1086-1099.

[9] C. C. Chen, and Y. Song. Solubility modeling with a non-random two-liquid segment activity coefficient model. Ind. Eng. Chem. Res. 2004, 43, 8354-8362.

[10] Van Winkle M. Effect of polar components on the relative volatility of the binary system n-hexane-benzene. J. Chem. Eng. Data 8 (1963) 210-214.

[11] H. Matsuda, K. Kaburagi, K. Kurihara, K. Tochigi, K. Tomono. Prediction of solubilities of pharmaceutical compounds in water + co-solvent systems using an activity coefficient model. Fluid Phase Equilibria 290 (2010) 153-157.

[12] D. Tassios. 62nd Annual Meeting, American Institute of Chemical Engineers, Washington, DC, November 1969.

[13] J. Gmehling and U. Onken. Vapor-Liquid Equilibrium Data Collection. DECHEMA, Franckfurt/Main, 1977.

[14] U. Weidlich, J. Gmehling. A modified UNIFAC model. 1. Prediction of VLE, hE. Ind. Eng. Chem. Res., 1987, 26 (7), pp 1372-1381.

[15] M. Sheikhzadeh, S. Rohani, M. Taffish, S. Murad. Solubility analysis of buspirone hydrochloride polymorphs: Measurements and prediction. Int. J. Pharm. 338 (2007) 55-63.

[16] S. Gracin, T. Brinck, and A. C. Rasmuson. Prediction of solid organic compounds in solvents by UNIFAC. Ind. Eng. Chem. Res. 2002, 41, 5114-5124.

[17] A. T. Kan and M. B. Tomson. UNIFAC prediction of aqueous and nonaqueous solubilities of chemicals with environmental interest. Environ. Sci. Technol. 1996, 30, 1369-1376.

[18] C. C. Chen and Y. Song. Extension of nonrandom two-liquid segment activity coefficient model for electrolytes. Ind. Eng. Chem. Res. 2005, 44, 8909-8921.

[19] Ehsan Sheikholeslamzadeh and Sohrab Rohani. Solubility prediction of pharmaceutical and chemical compounds in pure and mixed solvents using predictive models. Ind. Eng. Chem. Res. (2011) 51, 464-473.

[20] Ehsan Sheikholeslamzadeh and Sohrab Rohani. Vapour-liquid and vapour-liquid-liquid equilibrium modeling for binary, ternary, and quaternary systems of solvents. Fluid Phase Equilibria, 333, 97-105.

[21] Ehsan Sheikholeslamzadeh, Chau-Chyun Chen, and Sohrab Rohani. Optimal solvent screening for the crystallization of pharmaceutical compounds from multisolvent systems. Ind. Eng. Chem. Res. (2012) 51, 13792-13802.

[22] Ehsan Sheikholeslamzadeh and Sohrab Rohani. Modeling and optimal control of solution mediated polymorphic transformation of l-glutamic acid. Ind. Eng. Chem. Res. (2013) 52, 2633-2641.

[23] Stichlmair, J., Fair, J., Bravo, J. L. Separation of azeotropic mixtures via enhanced distillation. Chem. Eng. Prog. B 1989, 85, 63.

[24] Chen, R., Zhong, L., Xu, C. Isobaric vapor-liquid equilibrium for binary systems of toluene + ethanol and toluene + isopropanol at (1013.3, 121.3, 161.3, and 201.3) kPa. J. Chem. Eng. Data. 2011, 57, 155.

[25] Pequenín, A., Carlos Asensi, J., Gomis V. Isobaric vapor−liquid−liquid equilibrium and vapor−liquid equilibrium for the quaternary system water−ethanol−cyclohexane −isooctane at 101.3 kPa. J. Chem. Eng. Data, 2010, 55, 1227.

[26] Pequenín, A., Carlos Asensi, J., Gomis V. Experimental determination of quaternary and ternary isobaric vapor-liquid-liquid equilibrium and vapor-liquid equilibrium for the systems water-ethanol-hexane-toluene and water-hexane-toluene at 101.3 kPa. J. Chem. Eng. Data, 2010, 56, 3991.

[27] Kamihama, N., Matsuda, H., Kurihara, K., Tochigi, K., Oba, S. Isobaric vapor-liquid equilibria for ethanol + water + ethylene glycol and its constituent three binary systems. J. Chem. Eng. Data. dx.doi.org/10.1021/je2008704.

Chapter 2

Dynamics of Droplets

Hossein Yahyazadeh and Mofid Gorji-Bandpy

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/61901

Abstract

Capturing non-Newtonian power-law drops by horizontal thin fibers with circular cross-section in a quiescent media can be studied in this chapter. The case is simulated using volume of fluid (VOF) method providing a notable reduction of a computational cost. Open source OpenFOAM software is applied to conduct the simulations. This model is an extension of the one developed earlier by Lorenceau, Clanet, and Quéré [1]. To validate the model, water drops affecting a fiber of radius 350μm were simulated and threshold drop radiuses were obtained regarding to the impact velocity. These results agreed well with the experimental data presented by Lorenceau et al. [1]. In the next step, non-Newtonian power-law drops landing on thin fiber of radius 350μm were simulated. The final goal of this study was to obtain the threshold velocity and radius of a drop that is completely captured by the fiber. Threshold radiuses for both shear-thinning and shear-thickening drops were obtained and compared with corresponding Newtonian drops. Results show that the threshold radius of drop increases in a fixed velocity as n, power-law index, increases. Furthermore, shear-thinning nature of the drop leads to instabilities in high Reynolds numbers (Re) as it influences the fiber.

Keywords: Impact, Wetting, Filtration, Non-Newtonian Drop, Power-Law Model

1. Introduction

The problem of the removal of aerosol particles from gas streams has become of increasing importance from the standpoint of public health and the recovery of valuable products. Technology of controlling the aerosol particles or improving the liquid phase of aerosol is very important in many industrial processes such as oil and petroleum, electronic, mining, and food, as well as waste products like noxious emission of aerosol in chemical plants. There are several ways for this purpose among which fibrous filters are more popular so that it is obvious to try to improve their efficiency. The efficiency of collection and the pressure drop are the most important practical considerations in the design of these fibrous filters [2]. Various

experimental studies on liquid aerosol filtration have shown a filter clogging in several stages leading to a drainage stage with a constant pressure drop [3, 4]. Contal et al. [3] introduced four stages describing the clogging of fibrous filters by liquid droplets: In the first stage, the droplets impact the wet fibers. In the second stage, the amount of those droplets captured by fiber increases so that neighbor droplets coalesce. In the third stage, liquid shells form in the net that leads to a large increase of the pressure drop and to a dramatic decrease in the efficiency of the filter. Finally, there will be a pseudo-stationary state between the droplet collection and gravitational drainage of the liquid phase [1]. The first stage of clogging is studied in many researches experimentally. Patel et al. [5] investigated the drop breakup in a flow through fiber filters and the breakup probability for a drop. Hung and Yao [6] evaluated the impact of water droplets on a cylindrical object experimentally. Both these studies show that depending on the speed of the impacting droplets, the drops can be captured and broken into several pieces. Lorenceau et al. [1] investigated the threshold velocity and radius of drops captured by thin fibers. They have shown that the threshold impact velocity of drops decreases as the drop radius increases and vice versa. Following this study, Lorenceau et al. [7] investigated off-center impact of droplets on horizontal fibers.

Most of liquid droplets in aerosols produced in different industrial and natural processes show non-Newtonian behavior. Therefore, studying the phenomenon of impaction of a non-Newtonian droplet on thin fibers is of prime importance. All mentioned studies have been devoted to Newtonian fluids. Colliding of non-Newtonian droplets on thin fibers, to our knowledge, has not been investigated previously. However, colliding of non-Newtonian drops on flat plates is evaluated in some experimental and numerical studies. Haeri and Hashemabadi [8] studied spreading of power-law fluids on inclined plates both numerically and experimentally. Saïdi et al. [9] investigated experimentally the influence of yield stress on the fluid droplet impact control. Son and Kim [10] studied experimentally the spreading of inkjet droplet of non-Newtonian fluid on a solid surface with controlled contact angle at low Weber and Reynolds numbers. Kim and Baek [11] investigated the parameters that govern the impact dynamics of yield stress on the fluid droplets on solid surfaces.

In this chapter, we focus on the first stage of clogging and investigate the impaction of non-Newtonian power-law fluids on thin fibers. We aim to obtain the threshold radius of impacting droplets in different impact velocities. Effect of shear-thinning and shear-thickening behavior of droplets is evaluated and compared with corresponding Newtonian fluids. For this purpose, volume of fluid method is used and open source OpenFOAM software is applied for simulations.

The order of contents in this chapter is as follows: In Section 2 the governing equations and numerical methodology are given. Validation of the obtained results for Newtonian fluids in comparison with experimental observations presented by Lorenceau et al. [1] is provided in Section 3.1. A general description of the observations for non-Newtonian droplets captured by horizontal fiber is presented in Section 3.2, and the behavior of non-Newtonian drops affecting thin fibers is explained. Finally, a summary of the results and conclusions is provided.

http://dx.doi.org/10.5772/61901

2. Methodology

2.1. Volume of fluid method

Hirt and Nichols [12] demonstrated the volume of fluid (VOF) method and started a new trend in multiphase flow simulation. It relies on the definition of an indicator function γ. This function allows us to know whether one fluid or another occupies the cell, or a mix of both. In the conventional volume of fluid method [12], the transport equation for an indicator function γ, representing the volume fraction of one phase, is solved simultaneously with the continuity and momentum equations as follows:

$$\nabla.\mathbf{U} = 0 \tag{1}$$

$$\frac{\partial \gamma}{\partial t} + \nabla.\left(\mathbf{U}\gamma\right) = 0 \tag{2}$$

$$\frac{\partial\left(\rho\mathbf{U}\right)}{\partial t} + \nabla.\left(\rho\mathbf{U}\mathbf{U}\right) = -\nabla p + \nabla.\mathbf{T} + \rho\mathbf{f}_b \tag{3}$$

where $\mathbf{U}$ represents the velocity field shared by the two fluids throughout the flow domain, γ is the phase fraction, $\mathbf{T}=2\mu\,\mathbf{S} - 2\mu\,(\nabla.\mathbf{U})\,\mathbf{I}/3$ is the deviatoric viscous stress tensor, with the mean rate of strain tensor $\mathbf{S}=0.5[\nabla\mathbf{U} + (\nabla\mathbf{U})^T]$ and $\mathbf{I}=\delta_{i,j}$, ρ is density, p is pressure, and $\mathbf{f}_b$ are body forces per unit mass. In VOF simulations, the latter forces include gravity and surface tension effects at the interface. The phase fraction γ can take values within the range $0 \le \gamma \le 1$, with the values of zero and one corresponding to regions accommodating only one phase, for example, $\gamma=0$ for gas and $\gamma=1$ for liquid. Accordingly, gradients of the phase fraction are encountered only in the region of the interface. Two immiscible fluids are considered as one effective fluid throughout the domain, the physical properties of which are calculated as weighted averages based on the distribution of the liquid volume fraction, thus being equal to the properties of each fluid in their corresponding occupied regions and varying only across the interface:

$$\rho = \rho_1\,\gamma + \rho_g\left(1-\gamma\right) \tag{4}$$

$$\mu = \mu_1\,\gamma + \mu_g\left(1-\gamma\right) \tag{5}$$

where ρ_1 and ρ_g are densities of liquid and gas, respectively.

In this study, a modified approach similar to the one proposed in [13] is used.

2.2. Power-law model

Generally, the form of the function relating τ_{xy} to shear rate $\dot{\gamma}$ is quite complicated. It has been found, however, that the two-parameter equation of state

$$\tau_{xy} = K\dot{\gamma}^{n} \tag{6}$$

is adequate for many non-Newtonian fluids [16], where K is the fluid consistency coefficient and n is power-law index. This is a well-known power-law model, which is used in this study.

2.3. Computational method

All computations are performed using the code OpenFOAM [17], an open source computational fluid dynamics (CFD) toolbox, utilizing a cell-center-based finite volume method on a fixed unstructured numerical grid and employing the solution procedure based on the pressure implicit with splitting of operators (PISO) algorithm for coupling between pressure and velocity in transient flows [18].

A grid spacing equal to 1/20 of the droplet radius was used to discretize the computational domain. The mesh size was determined based on a mesh refinement study in which the grid spacing was progressively decreased until further reductions made no significant change in the predicted droplet shape during the impact. Bussmann et al. [19] have provided a detailed description of such a mesh refinement study earlier. The mesh size was uniform throughout the computational domain, which included the droplet and fiber.

3. Results and discussion

3.1. Validation of the Solution

In this section, we present the results obtained for water droplets impacting a fiber of radius 350 μm. Physical properties of water are given in Table 1. Threshold radiuses of the droplets in different impact velocities are obtained and compared with those exhibited by Lorenceau et al. [1]. All obtained results are shown in Figure 1, which demonstrates the threshold radiuses in different impact velocities.

n	$\rho\ (kg/m^3)$	$k\ (mPa\cdot s^n)$	$\sigma\ (mN/m)$	fluid
1	1000	1	70	water

Table 1. Physical properties of fluids

http://dx.doi.org/10.5772/61901

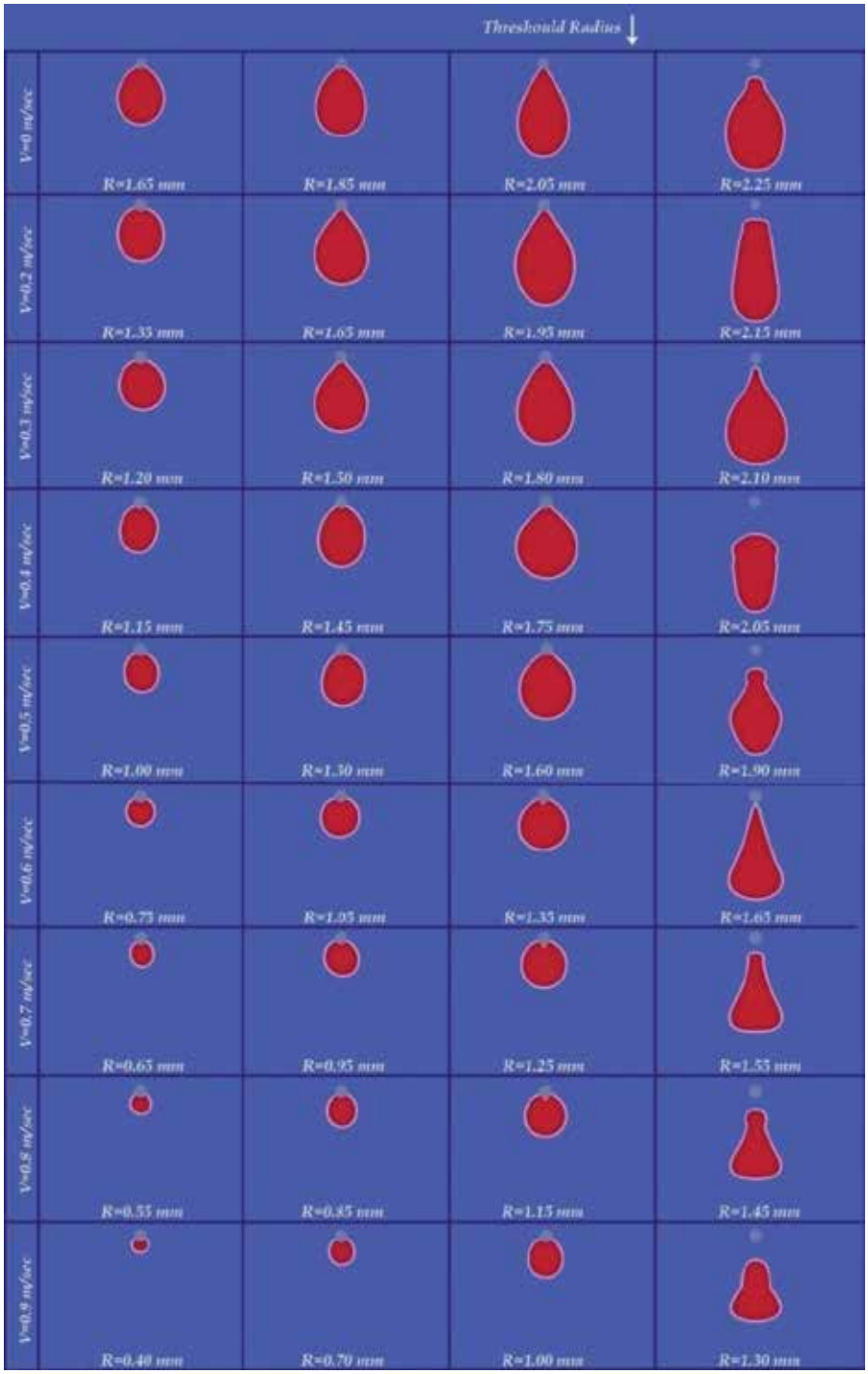

Figure 1. Threshold radiuses of the water droplets in different impact velocities

Defining R_M as characteristic radius of the drop at zero impact velocity as follows:

$$R_M = 3^{1/3} b^{1/3} k^{-2/3} \tag{7}$$

where $k^{-1}=\sqrt{\sigma/\rho g}$ is the capillary length and b is the radius of the fiber, and also characteristic velocity of the drop as:

$$U_M = \sqrt{4gR_M} \tag{8}$$

Variation of the dimensionless threshold radius versus dimensionless impact velocity is plotted in Figure 2, which shows a good agreement with the experimental and theoretical data presented by Lorenceau et al. [1].

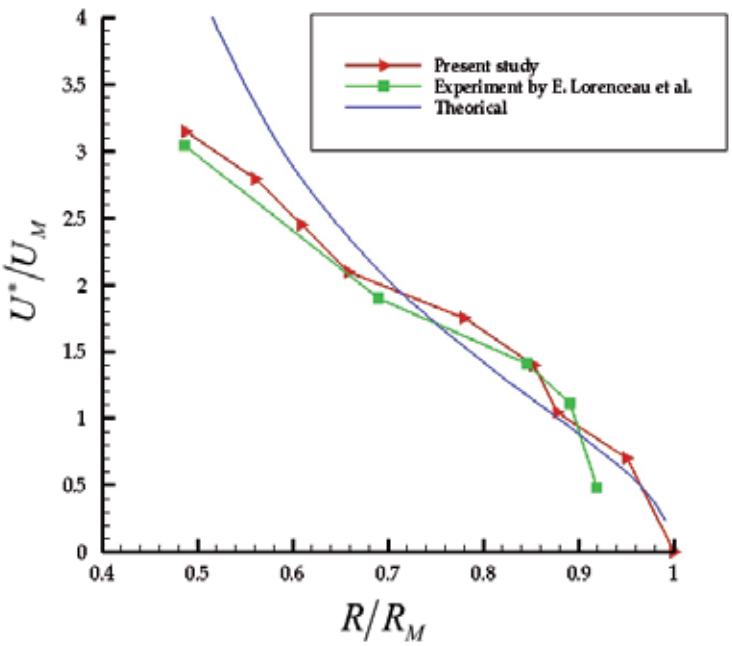

Figure 2. Comparison of the obtained results with experimental and theoretical data

Errors of the achieved results in comparison with experimental results of Lorenceau et al. (2004) are presented in Table 2.

% Errors in comparison with theoretical data	% Errors in comparison with experimental data	U/UM Theoretical data	U/UM Experimental data	U/U_M Present study	R/R_M
3.68	12.6	1.00	1.19	1.04	0.87
23	2.96	1.13	1.35	1.39	0.85
14.47	8.07	1.52	1.65	1.74	0.78
11.06	0.48	2.35	2.08	2.09	0.65
12.54	4.27	2.79	2.34	2.44	0.60
16.46	6.48	3.34	2.62	2.79	0.56
30.53	3.63	4.52	3.03	3.14	0.48

Table 2. Errors of the achieved results in comparison with the experimental and theoretical data

3.2. Results for non-newtonian droplets

Four forces, capillary, gravity, inertia, and viscosity, are important during the impaction of droplet to the fiber. Capillary and gravity forces are in contrast. While the weight of the drop tends to detach it from the fiber, surface tension is responsible for the sustentation. In order to analyze the dynamics of the capture of a drop by a fiber, viscous and inertia effects must be considered. Here again, those effects are antagonistic. While inertia tends to make the drop cross the fiber, viscous effects will induce a dissipation and can thus possibly stop the drop [1]. Effects of variation of viscosity were not investigated in the previous works. In order to investigate these effects, both shear-thinning and shear-thickening fluids will be considered.

At first, we present general observations of threshold radiuses at different velocities for shear-thinning droplets, $n<1$. For this purpose, physical properties of the fluid given in Table 1 were used except the power-law index, which is equal to 0.5. Figure 3 represents the threshold radiuses in different velocities for this type of droplet.

Next, we present our observations for the two kinds of shear-thickening droplets, n=1.5, 2, impacting the fiber which are shown in Figures 4 and 5.

Results show that, on the one hand, in shear-thinning drops, the threshold radius decreased in a fixed velocity in comparison with the corresponding Newtonian drops. Moreover, instabilities were observed during the impaction of the drop. This is because viscosity in shear-thinning fluids decreases as it has an inverse relation with shear rate. That is, as a shear-thinning droplet impacts a thin fiber, due to high shear rates, viscosity decreases in comparison with the corresponding Newtonian fluid, so that the viscosity force reduces (i.e., a force that helps the drop to stick on the fiber is reduced). Thus, in order to balance the antagonistic forces, drop size has to be decreased, which means the threshold radius has to be decreased. On the other hand, in the case where the drop is a shear-thickening fluid, the threshold radius increased in a fixed velocity in comparison with the corresponding Newtonian fluid, since the viscosity of shear-thickening fluids has a direct relation to the shear rate. It means that, viscosity increases as the drop impacts the fiber compared to Newtonian drops. Thereby, in shear-

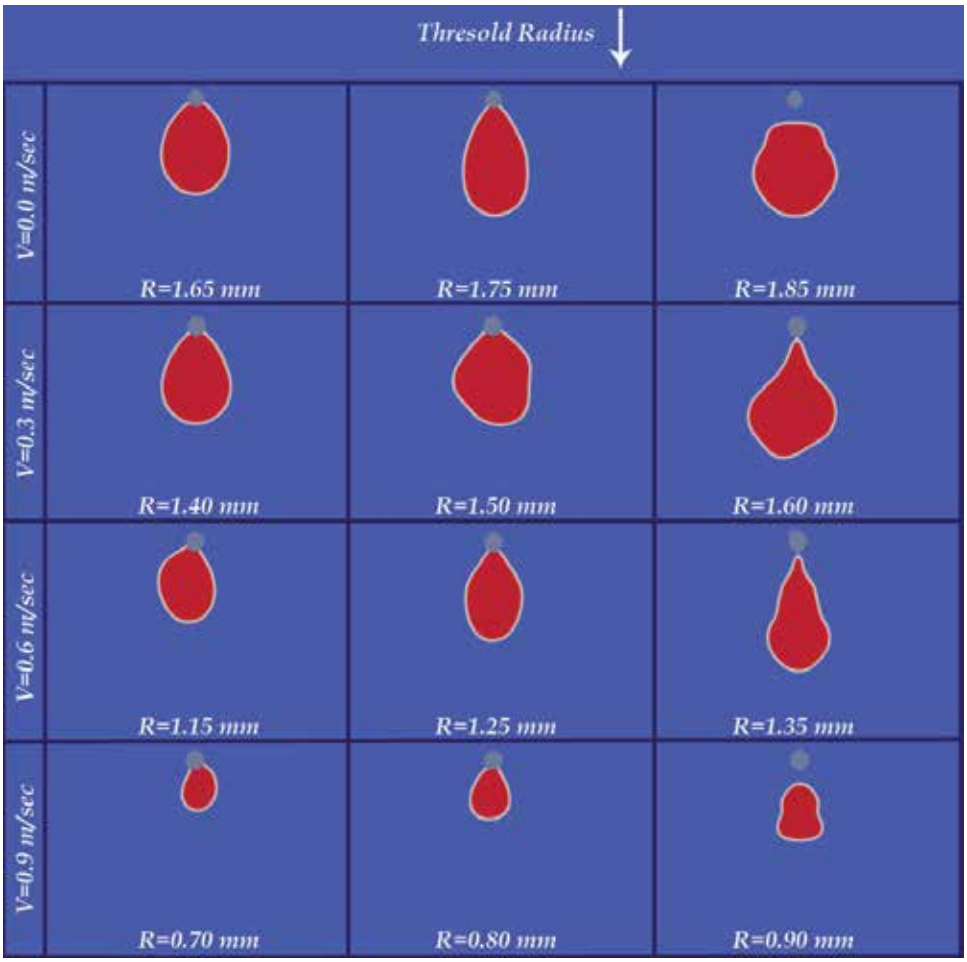

Figure 3. Threshold radiuses of the shear-thinning (n=0.5) droplets in different impact velocities

thickening drops, viscous forces increase and the drop's tendency to stick to the fiber increases. Therefore, the threshold drop size will also increase. This tendency is obvious in Figures 4 and 5. Threshold radius of Newtonian, shear-thinning and shear-thickening droplets versus impact velocity is plotted in Figure 6.

Figure 6 shows that the threshold radius will decrease generally with the increase of an impact velocity or vice versa. In shear-thinning droplets, viscosity has an inverse relation with shear rate, as mentioned before, so that it will decrease with the increase of impact velocity. It means inertia forces increase whereas viscosity forces decrease. That is why instabilities are observed in shear-thinning droplets impacting fiber at high speeds. The threshold radius of shear-thickening droplets, also, decreases with the increase of impact velocity. In shear-thickening droplets, although there is a direct relation between the viscosity and the shear rate, velocity increases with second power compared to capillary forces. That is, increase of inertia forces is

http://dx.doi.org/10.5772/61901

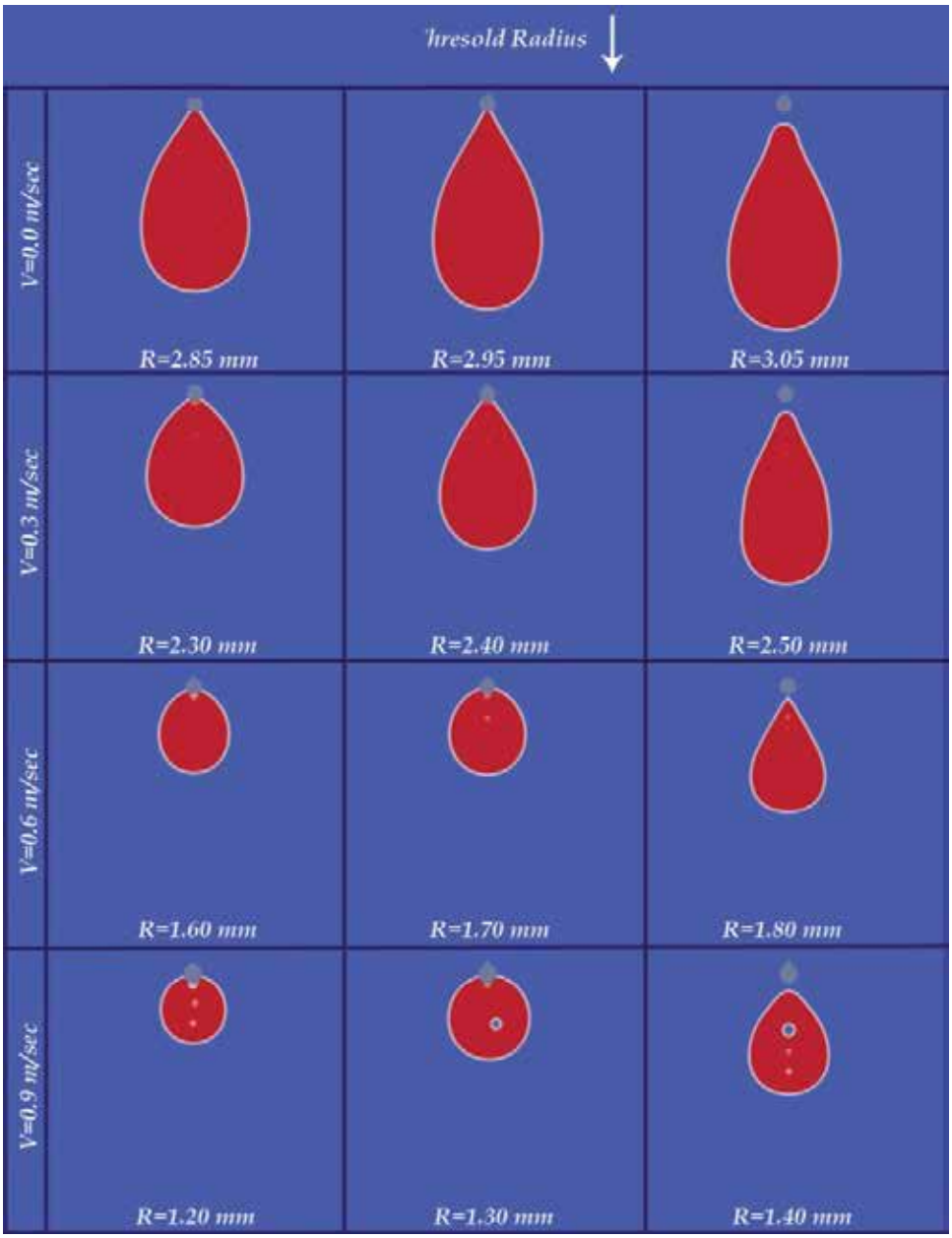

Figure 4. Threshold radiuses of the shear-thickening (n=1.5) droplets in different impact velocities

higher than increase of viscosity forces so that the threshold radius will decrease to make the drop stick to the fiber. However, in higher velocities, this rate will decrease as it can be seen in Figure 6. In other words, rate of the decreasing of threshold radius with the increasing of impact velocity will decrease due to high growth of shear rate.

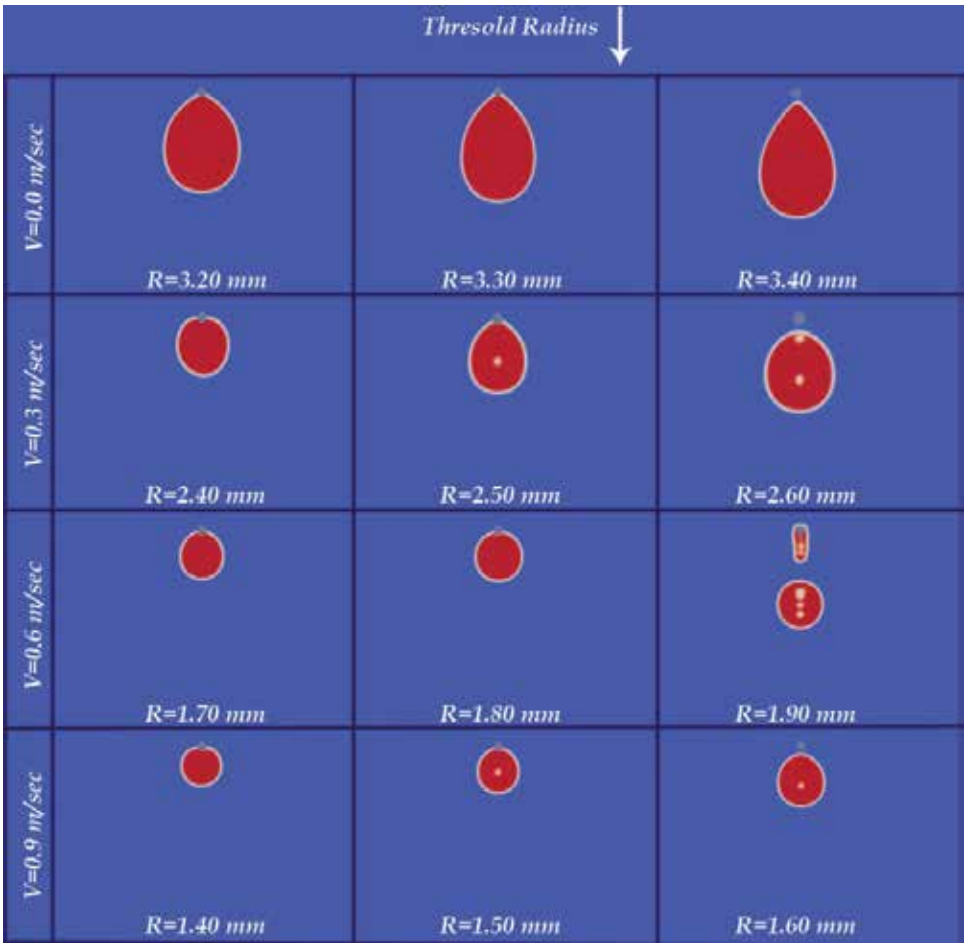

Figure 5. Threshold radiuses of the shear-thickening (n=2) droplets in different impact velocities

There are some general observations that are common to Newtonian and non-Newtonian fluids. For all kinds of fluids, the threshold radius of the droplet has a reverse relation with the impact velocity of the droplet. In all cases, at a fixed impact velocity, droplets with a radius greater than the threshold radius have passed the fiber without breakup, and drops with a radius lower than the threshold radius clung to the fiber entirely.

4. Conclusion

Impaction of non-Newtonian power-law droplet to the horizontal fiber of circular cross section is investigated in this study. Volume of fluid technique is employed, significantly reducing the computational cost. Outcomes are divided into three parts: First, it has been observed that the threshold radius of the droplets decreased with the increase of impact velocity for New-

http://dx.doi.org/10.5772/61901

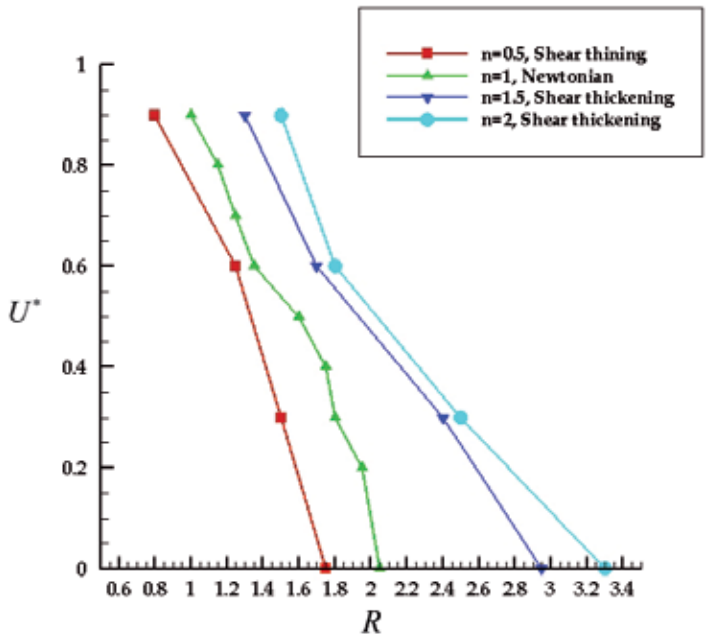

Figure 6. Threshold radiuses of droplets versus impact velocity for Newtonian, shear-thinning and shear-thickening fluids

tonian, shear-thinning and shear-thickening fluids, due to the growth of inertia forces versus viscosity and capillary forces. Second, instabilities have been observed in shear-thinning droplets at high impact velocities due to severe reduction of viscosity. Finally, in shear-thickening droplets, at high impact speeds, rate of reduction of threshold radius has decreased due to increase of the viscosity forces.

Author details

Hossein Yahyazadeh and Mofid Gorji-Bandpy*

*Address all correspondence to: gorji@nit.ac.ir

Department of Mechanical Engineering, Babol University of Technology, Babol, Iran

References

[1] É. Lorenceau, C. Clanet, D. Quéré, Capturing drops with a thin fiber, Journal of Colloid and Interface Science, vol. 279, 192–197, 2004.

[2] C.Y. Chenz, Filtration Of Aerosols By Fibrous Media,DA18-108-CML-4789 with the Chemical Corps, U.S. Army, Washington 25, D.C, March 6, 1955.

[3] P. Contal, J. Simao, D. Thomas, T. Frising, S. Call, J.C. Appert-Collin, D. Bemer, Clogging of fibre filters by submicron droplets. Phenomena and influence of operating conditions, Aerosol Science, vol. 35, 263–278, 2004.

[4] D.C. Walsh, J.I.T. Stenhouse, K.L. Scurrah, A. Graef, The effect of solid and liquid aerosol particle loading on fibrous filter material performance, Journal of Aerosol Science, vol. 27, 617–618, 1996.

[5] P. Patel, E. Shaqfeh, J.E. Butler, V. Cristini, J. Blawzdziewicz, M. Loewenberg, Drop breakup in the flow through fixed fiber beds: An experimental and computational investigation, Physics of Fluids, vol. 15, 1146, 2003.

[6] L.S. Hung, S.C. Yao, Experimental investigation of the impaction of water droplets on cylindrical objects, International Journal of Multiphase Flow, vol. 25, 1545–1559, 1999.

[7] E. Lorenceau, C. Clanet, D. Quéré, M. Vignes-Adler, Off-centre impact on a horizontal fibre, European Physical Journal Special Topics, vol. 166, 3–6, 2009.

[8] S. Haeri, S.H. Hashemabadi, Three dimensional CFD simulation and experimental study of power law fluid spreading on inclined plates, International Communications in Heat and Mass Transfer, vol. 35, 1041–1047, 2008.

[9] Alireza Saïdi, Céline Martin, Albert Magnin, Influence of yield stress on the fluid droplet impact control, Journal of Non-Newtonian Fluid Mechanics, vol. 165, Issues 11–12, 596–606, 2010.

[10] Yangsoo Son, Chongyoup Kim, Spreading of inkjet droplet of non-Newtonian fluid on solid surface with controlled contact angle at low Weber and Reynolds numbers, Journal of Non-Newtonian Fluid Mechanics, vol. 162, Issues 1–3, 78–87, 2009.

[11] Eunjeong Kim, Jehyun Baek, Numerical study of the parameters governing the impact dynamics of yield-stress fluid droplets on a solid surface, Journal of Non-Newtonian Fluid Mechanics, vol. 173–174, 62–71, 2012.

[12] C.W. Hirt, B.D. Nichols, Volume of fluid (VOF) method for the dynamics of free boundaries, Journal of Computational Physics, vol. 39, 201–225, 1981.

[13] H. Rusche, Computational Fluid Dynamics of Dispersed Two-Phase Flows at High Phase Fractions, Imperial College of Science, Technology and Medicine, Ph.D. thesis, 2002.

[14] G. Černe, S. Petelin, I. Tiselj, Coupling of the interface tracking and the two-fluid models for the simulation of incompressible two-phase flow, Journal of Computational Physics, vol. 171, 776–804, 2001.

[15] J.U. Brackbill, D. B. Kothe, C. Zemach, A continuum method for modeling surface tension, Journal of Computational Physics, vol. 100, 335–354, 1992.

[16] Rohit Aiyalur Shankaran, Numerical Simulation Of Flow Of Shear-Thinning Fluids in Corrugated Channels, M.Sc. thesis, Texas A&M University, December 2007.

[17] H.G. Weller, G. Tabor, H. Jasak, C. Fureby, A tensorial approach to computational continuum mechanics using object orientated techniques, Computers in Physics, vol. 12, 620–631, 1998.

[18] R.I. Issa, Solution of the implicitly discretised fluid flow equations by operator-splitting, Journal of Computational Physics, vol. 62, 40–65, 1986.

[19] M. Bussmann, J. Mostaghimi, S. Chandra, On a three-dimensional volume tracking model of droplet impact, Physics of Fluids, vol. 11, 1406–1417, 1999.

Chapter 3

Nonequilibrium Thermodynamic and Quantum Model of a Damped Oscillator

Gyula Vincze and Andras Szasz

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/61010

Abstract

We describe the linearly damped harmonic quantum oscillator in Heisenberg's interpretation by Onsager's thermodynamic equations. Ehrenfest's theorem is also discussed in this framework. We have also shown that the quantum mechanics of the dissipative processes exponentially decay to classical statistical theory.

Keywords: quantum oscillator, damping, Onsager's theory, Ehrenfest's theory, classic limit

1. Introduction

Dissipation is essential for the evolution of a quantum-damped oscillator. It is responsible for the decay of quantum states, the broadening of the spectral line, and the shifting the resonance frequency. This has been a persistent challenge for a long time since dissipation causes difficulties in the quantization of the damped oscillator [1, 2, 3]. This problem has remained under intensive investigation [4, 5]. There are some widely accepted Hamilton-like variation theories about the treatment of a linearly damped classic or quantum-damped oscillator. One of these theories is Bateman's mirror-image model [1], which consists of two different damped oscillators, where one of them represents the main linearly damped oscillator. The energy dissipated by the main oscillator will be absorbed by the other amplified oscillator, and thus the energy of the total system will be conserved. The fundamental commutation relations of this model are time independent; however, the time-dependent uncertainty products, obtained in this way, vanish as time tends to infinity [6]. The Caldirola–Kanai theory with an explicit

time-dependent Hamiltonian is another kind of variation theory [7, 8, 9]. In the quantum version of this theory, both the canonical commutation rules and the uncertainty products tend to zero as time tends to infinity. The system-plus-reservoir model [10, 11] is another damped oscillator model. It is coupled linearly to a fluctuating bath. If the bath is weakly perturbed by the system, then it can be modeled with a continuous bath of the harmonic oscillator. A quantum Langevin equation in the form of a Heisenberg operator differential equation can be deduced in this model. However, this equation in general does not obey Onsager's regression hypothesis [12], i.e., only in case when $\hbar \to 0$ [13]. A direct consequence of this fact is that the expected value of the fundamental observable does not satisfy the equation of the classic linearly damped oscillator. Another consequence is that no spontaneous dissipative process exists in this theory. The above models are based on the Heisenberg's mechanical reinterpretation model [14].

A possible reinterpretation model based on irreversible thermodynamics was recently published [15]. This model started from the Rosen–Chambers restricted variation principle of the nonequilibrium thermodynamics [16, 17, 18] and used a Hamilton-like variation approach to the linearly damped oscillator. The usual formalisms of classical mechanics, such as the Lagrangian, Hamiltonian, and Poisson brackets, were also covered by this variational principle. By means of canonical quantization, the quantum mechanical equations of the linearly damped oscillator are given. The resulting Heisenberg operator differential equations of the damped oscillator are consistent with the classical equations of motion and can be solved by using ladder operators, which are time dependent. By this theory, the exponential decay of quantum states, the natural width of the spectral line, and the shifts in the resonance frequency can be explained. This work describes the quantum theory of a linearly damped oscillator, which could be reinterpreted in terms of a classical model based on Onsager's nonequilibrium thermodynamic theory, corresponding to the Heisenberg reinterpretation principle. The first chapters are devoted to Onsager's thermodynamic theory and the quantum theory of a damped oscillator. The dissipative quantum theory given in the Heisenberg picture is deduced from the general evolution equation of a Hermitian observable by means of two system-specific constitutive equations. The first of the constitutive equations belongs to unitary dynamics, while the second belongs to the dissipative dynamics of the observable. The fundamental commutators, which are a consequence of the constitutive equations, are time dependent. The quantum mechanical equations of motion of the oscillator in the Heisenberg picture, the Ehrenfest theorem, and the uncertainty principle of that oscillator are given. A significant part of the work deals with applications such as the expected value of the main operators of the damped oscillator, the probability description of the wave packet motion belonging to the damped oscillator, the calculation of the wave function by matrix calculus, the spectral density of the energy dissipation, and the natural width of the spectral line. Another significant part of this work deals the quantum statistics of the damped oscillator. By a generalization of the Liouville–von Neumann equation, the statistical thermodynamic theory of the ensemble of the damped oscillators in contact with a thermal bath is given. By introducing the quantum entropy of the ensemble, it is shown that the entropy of the ensemble grows in a dissipative process and in thermal equilibrium for the probability distribution of the quantum states, such that Gibbs' canonical distribution is valid. Finally, a wave equation of the linearly damped oscillator is given.

http://dx.doi.org/10.5772/61010

2. Nonequilibrium thermodynamic theory of the linearly damped oscillator

Meixner was the first to propose a nonequilibrium thermodynamic theory for linear dissipative networks [19, 20]; for a general overview on network thermodynamics, see [21]. In this theory, it can be shown that, for example, electrical networks are thermodynamic systems, and it is possible to derive the network equations (Kirchhoff equations) by application of the principles of nonequilibrium thermodynamics. In what follows, we give an Onsagerian thermodynamic theory of the linearly damped harmonic oscillator. A damped oscillator, as a primitive network, is considered under the isotherm condition, which is maintained by removing the irreversible heat as it developed in damping resistance, or in other words, by placing the damping resistance of the oscillator in a temperature bath. In this case, it is possible to speak not only of the entropy of the damped oscillator but also of the free energy, and not of entropy production but of energy dissipation instead. To show these, we give the actual form of the first law of thermodynamics in the case of a damped oscillator. To do this, let us introduce from the energy conservation law of the oscillator + thermal bath system (Figure 1), such that we obtain the following:

$$\frac{d}{dt}\left(U + U_{bath}\right) = 0 \rightarrow \frac{dU}{dt} = -\frac{dU_{bath}}{dt} \tag{1}$$

where U and U_{bath} are the internal energies of the oscillator and the thermal bath. Since the oscillator is isolated by a rigid diathermal wall, the energy exchange between the oscillator and the bath must be heat transfer only. Thus, the first law of thermodynamics of the oscillator and the thermal bath has the following forms

$$\frac{dU}{dt} = \frac{dQ}{dt}, \quad -\frac{dU_{bath}}{dt} = \frac{dQ}{dt} \tag{2}$$

where $\frac{dQ}{dt}$ is the power of exchanged heat on the rigid diathermal wall.

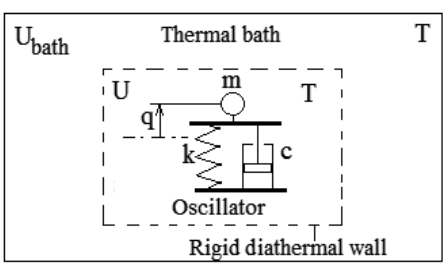

Figure 1. Mechanical equivalent circuit of a linearly damped oscillator.

Assume that the entropy S of the oscillator exists, and it is a state function of the variables U, q. By the second law of nonequilibrium thermodynamics, the entropy $S(U, q)$ satisfies the balanced equation:

$$\frac{dS}{dt} = \frac{dS_r}{dt} + \frac{dS_i}{dt}, \quad \frac{dS_i}{dt} \geq 0 \tag{3}$$

where $\frac{dS_r}{dt}$ is the reversible rate of change of the entropy or entropy flux on the rigid diathermal wall and $\frac{dS_i}{dt}$ is the so-called entropy production. From the second law of the thermodynamics follows the well-known expression for entropy flux:

$$\frac{dS_r}{dt} = \frac{1}{T}\frac{dQ}{dt} \tag{4}$$

On the other hand, the rate of change in the entropy of the oscillator can be written as

$$\frac{dS}{dt} = \frac{\partial S}{\partial U}\frac{dU}{dt} + \frac{\partial S}{\partial q}\frac{dq}{dt} \tag{5}$$

From this equation and Equations (2) and (4), the following relations can result:

$$\frac{\partial S}{\partial U} = \frac{1}{T}, \quad \frac{dS_i}{dt} = \frac{\partial S}{\partial q}\frac{dq}{dt} \geq 0 \tag{6}$$

Thus, the entropy balance equation has the form

$$\frac{dS}{dt} = \frac{1}{T}\frac{dQ}{dt} + \frac{\partial S}{\partial q}\frac{dq}{dt} \tag{7}$$

Because the temperature of the oscillator is constant, we could introduce

$$F = U - TS \tag{8}$$

i.e., the free energy of the oscillator, and from Equations (2) and (7), we can obtain the simple balanced equation for free energy

$$\frac{dF}{dt} = -T\frac{\partial S}{\partial q}\frac{dq}{dt} \leq 0 \tag{9}$$

http://dx.doi.org/10.5772/61010

Also, entropy production and the so-called rate of energy dissipation

$$R = T\frac{dS_i}{dt} = T\frac{\partial S}{\partial q}\frac{dq}{dt} \geq 0 \tag{10}$$

decrease the free energy of the oscillator. In what follows, we shall give the actual form of the rate of energy dissipation. Next, we see from Equation (10) that the rate of energy dissipation must be some explicit function of $\frac{dq}{dt}$ and may depend implicitly on U, q and the following equation

$$R = R(\frac{dq}{dt}; U, q) \geq 0 \tag{11}$$

such that

$$R(0; U, q) = 0 \tag{12}$$

is therefore quite general. We now expand the energy dissipation Equation (11) in a Taylor series, i.e.,

$$R = A_0 + A_1\frac{dq}{dt} + \frac{1}{2}A_2\left(\frac{dq}{dt}\right)^2 + ... \tag{13}$$

The sufficient condition of nonnegativity of entropy production, which will always be satisfied, is the complete exclusion of all odd terms in $\frac{dq}{dt}$ in Equation (13), and A_0 must be zero to exclude entropy production in an equilibrium state. Thus, one need only neglect the fourth-order term in the Taylor series (Equation (13)) to obtain the Rayleigh dissipation function

$$R = c\left(\frac{dq}{dt}\right)^2, \quad c = \frac{1}{2}A_2 \tag{14}$$

In addition, the so-called damping constant of oscillator c must be positive to satisfy the nonnegativity condition in Equation (13). If we assume that the nondissipative elements of the damped oscillator are linear, then the free energy in Equation (8) can be identified with the energy stored in the mass m and the spring of the oscillator (Figure 1.). Also, the free energy is equal to the Hamiltonian of the oscillator, i.e.,

$$F = H = \frac{1}{2}\frac{p^2}{m} + \frac{1}{2}kq^2 \tag{15}$$

where q is the displacement of the mass from its equilibrium position, p is the momentum of the mass of the oscillator, and k is the constant of the spring. A direct consequence of the above results is that R can be constructed as a bilinear form, namely,

$$R = -\frac{dH}{dt} = \frac{dq}{dt}\left(-\frac{\partial H}{\partial q}\right) + \frac{dp}{dt}\left(-\frac{\partial H}{\partial p}\right) \geq 0 \tag{16}$$

Here we now interpret, in the usual nonequilibrium thermodynamic fashion, the quantities $\left(\frac{dq}{dt}, \frac{dp}{dt}\right)$ in terms of thermodynamic fluxes and the quantities $\left(-\frac{\partial H}{\partial q}, -\frac{\partial H}{\partial p}\right)$ in terms of thermodynamic forces. Since we now have Equation (14) of the rate of energy dissipation of the damped oscillator, Equation (16) can be written in the form

$$\frac{dq}{dt}\left(-\frac{\partial H}{\partial q}\right) + \frac{dp}{dt}\left(-\frac{\partial H}{\partial p}\right) = c\left(\frac{dq}{dt}\right)^2 \geq 0 \tag{17}$$

In Onsagerian thermodynamics, the constitutive (kinetic) equations between fluxes and forces are linear

$$\begin{bmatrix} \frac{dq}{dt} \\ \frac{dp}{dt} \end{bmatrix} = \begin{bmatrix} L_{qq} & L_{qp} \\ L_{pq} & L_{pp} \end{bmatrix} \begin{bmatrix} -\frac{\partial H}{\partial q} \\ -\frac{\partial H}{\partial p} \end{bmatrix} \tag{18}$$

where the kinetic matrix

$$\mathbf{L} = \begin{bmatrix} L_{qq} & L_{qp} \\ L_{pq} & L_{pp} \end{bmatrix} \tag{19}$$

can be split into nondissipative (so-called reactive) and dissipative parts. To do this, take these kinetic equations into Equation (17), and using Equation (15), with a simple calculation, we obtain for these kinetic matrices

http://dx.doi.org/10.5772/61010

$$\begin{bmatrix} \frac{dq}{dt} \\ \frac{dp}{dt} \end{bmatrix} = \left(\begin{bmatrix} 0 & -a \\ a & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & c \end{bmatrix} \right) \begin{bmatrix} -\frac{\partial H}{\partial q} \\ -\frac{\partial H}{\partial p} \end{bmatrix} \tag{20}$$

Here, a is an arbitrary constant. Now, we see that the dissipative part of the kinetic matrix satisfies the Onsager symmetry relation and the positivity of the damping constant c trivially. The constant a can be evaluated as follows. In the case of zero for the dissipative part of the kinetic matrix, these equations must be transformed into Hamilton equations of a simple harmonic oscillator. From this fact, it follows that a is a universal constant and $a=1$. Also, the final form of the Onsagerian constitutive (kinetic) equation the of damped oscillator is

$$\begin{bmatrix} \frac{dq}{dt} \\ \frac{dp}{dt} \end{bmatrix} = \left(\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & c \end{bmatrix} \right) \begin{bmatrix} -\frac{\partial H}{\partial q} \\ -\frac{\partial H}{\partial p} \end{bmatrix} \tag{21}$$

It is easy to show that these kinetic equations are equivalent to the Newtonian equations of motion of the linearly damped oscillator, namely,

$$\frac{dq}{dt} = \frac{\partial H}{\partial p} = \frac{p}{m}, \quad \frac{dp}{dt} = -\frac{\partial H}{\partial q} - c\frac{\partial H}{\partial p} = -kq - c\frac{p}{m} \tag{22}$$

From this equivalence, it follows that the velocity of oscillator as a generalized thermodynamic flux has only a reactive part, while the rate of momentum, as another thermodynamic flux, has both reactive and dissipative constituents. The presented thermodynamic deduction of equations of a linearly damped oscillator enables us to build a stochastic force F_s into equations of motion (Equation (22)). This stochastic force and the thermodynamic forces introduced above are statistically independent and take into account the effect of the temperature bath. In this case, the thermodynamic fluxes are fluctuations, which obey another type of equation (Equation (22)). The correct form of these equations follows from the regression hypothesis of Onsager [22], which states that "the average regression of fluctuations will obey the same laws as the corresponding macroscopic irreversible process." From this, we give the Langevin-type equations for a linearly damped oscillator,

$$\frac{dq}{dt} = \frac{\partial H}{\partial p}, \quad \frac{dp}{dt} = -\frac{\partial H}{\partial q} - c\frac{\partial H}{\partial p} + F_s \tag{23}$$

A consequence of these equations is that the dissipative kinetic coefficient c can be related to the correlation coefficient $\langle p(t),\ p(t+t')\rangle$ via the fluctuation dissipation theorem [23, 24]. According to Equation (22), we can conclude that the fluctuations of thermodynamic fluxes are similar to an impressed macroscopic deviation, except they appear spontaneously.

2.1. Bohlin's first integral

By means of Equation (21) or (22), we can deduce the time rate of change of any observable defined in the phase space of the damped oscillator. Let $O(q,\ p,\ t)$ be an observable, such that its time rate of change can be expressed as

$$\frac{dO}{dt}=\frac{\partial O}{\partial t}+\frac{\partial O}{dq}\frac{dq}{\partial t}+\frac{\partial O}{\partial p}\frac{dp}{dt}=\frac{\partial O}{\partial t}+\frac{\partial O}{dq}\frac{\partial H}{\partial p}-\frac{\partial O}{\partial p}\frac{\partial H}{\partial q}-c\frac{\partial O}{\partial p}\frac{\partial H}{\partial p} \tag{24}$$

where we take into account the Onsagerian equations (Equation (22)). An observable $O(q,\ p,\ t)$ is the first integral of the (Onsagerian equation (Equation (21)) of the damped oscillator if

$$\frac{dO}{dt}=0 \tag{25}$$

and if it is a constant of the motion. Now, let us give the constant of the motion of the linearly damped harmonic oscillator. It was Bohlin [25] who first dealt with the problem of the constants of motion for a damped linear oscillator. It is easy to prove that Bolin's observable, defined as

$$\begin{aligned}B\left(q,p,t\right)&=\frac{m}{2}e^{2\beta t}\left(\frac{dq}{dt}-\gamma q\right)\left(\frac{dq}{dt}-\gamma^{*}q\right)=\\&=e^{2\beta t}\left(\frac{p^{2}}{2m}+\frac{\beta}{2}qp+\frac{\beta}{2}pq+\frac{m\omega_{0}^{2}}{2}q^{2}\right)\end{aligned} \tag{26}$$

where

$$\beta=\frac{c}{2m},\omega_{0}=\sqrt{\frac{k}{m}},\gamma=-\beta+i\omega,\qquad\gamma^{*}=-\beta-i\omega,\quad\omega=\sqrt{\omega_{0}^{2}-\beta^{2}} \tag{27}$$

is the first integral of the damped oscillator. Now, we may see that the Bolin's observable in the case of the undamped oscillator is equal to the Hamiltonian of the oscillator.

http://dx.doi.org/10.5772/61010

3. Quantum theory of linearly damped oscillator

In the standard theory of quantum mechanics, two kinds of evolution processes are introduced, which are qualitatively different from each other. One is the spontaneous process, which is a reactive (unitary) dynamical process and is described by the Heisenberg or Schrödinger equation in an equivalent manner. The other is the measurement process, which is irreversible and described by the von Neumann projection postulate [26], which is the rigorous mathematical form of the reduction of the wave packet principle. The former process is deterministic and is uniquely described, while the latter process is essentially probabilistic and implies the statistical nature of quantum mechanics.

3.1. The general evolution equation of the Hermitian operator

Unlike classical quantum mechanics, the spontaneous processes of the damped oscillator are irreversible, so its quantum mechanical description needs changes to some instruments of classical quantum mechanics. To do this, we use the Heisenberg picture of quantum processes. In this picture, the observables are time-dependent linear Hermitian operators, and the state vector of the system is time independent. Using the terminology introduced in the first part, the infinitesimal time transformation of the Hermitian operator could happen in two ways:

- By reactive transformation, when the orthonormal eigenvectors of the Hermitian observable turn in time, keeping the orthonormal system with unchanged eigenvalues. The eigenvectors belonging to the different moments are connected with unitary transformation, as in classic quantum mechanics. Dynamics belonging to this transformation are so-called unitary dynamics.
- By dissipative transformation, when the real eigenvalues of the Hermitian operator change irreversibly in time.

Let us study the general evolution equation of the Hermitian operator, considering both the above time-dependent processes. For simplicity in demonstrating the derivation, we suppose a discrete eigenvalue spectrum of the Hermitian operator, although the spectra of the displacement and momentum operators could be continuous. In this case, the orthonormal eigenvectors $|\Psi_{Oi}(t)\rangle$ and eigenvalues $\lambda_{Oi}(t)$ are solutions of the eigenvalue equation

$$\mathbf{O}(t)\left|\Psi_{Oi}(t)\right\rangle = \lambda_{Oi}(t)\left|\Psi_{Oi}(t)\right\rangle, \quad (i = 1,2,3,...) \tag{28}$$

and the eigenvectors form a complete orthonormal basis in a Hilbert space, when the eigenvalue spectrum is nondegenerate. Thus, the spectral representation of the operator would be

$$\mathbf{O}(t) = \mathbf{O}(t)\boldsymbol{\delta} = \sum_{i\geq 0}\lambda_{Oi}(t)\left|\Psi_{Oi}(t)\right\rangle\left\langle\Psi_{0i}(t)\right|, \quad \boldsymbol{\delta} = \sum_{i\geq 0}\left|\Psi_{Oi}(t)\right\rangle\left\langle\Psi_{0i}(t)\right| \tag{29}$$

where δ is the unity operator and $\langle\Psi_{Oi}(t)|$ is the dual of $|\Psi_{Oi}(t)\rangle$, so $\langle\Psi_{Oi}(t)|\Psi_{Oi}(t)\rangle$=1.

Since the observable is Hermitian, the transformation of the eigenvector $|\Psi_{Oi}(0)\rangle$ at t=0 to the eigenvector $|\Psi_{Oi}(t)\rangle$ at time t will be represented by an unitary operator $\mathbf{U}(t)$. Hence,

$$\left|\Psi_{Oi}(t)\right\rangle = \mathbf{U}(t)\left|\Psi_{Oi}(0)\right\rangle, \quad \left\langle\Psi_{Oi}(t)\right| = \mathbf{U}^{+}(t)\left\langle\Psi_{Oi}(0)\right|, \quad \mathbf{U}(t)\mathbf{U}^{+}(t) = \boldsymbol{\delta} \tag{30}$$

Let us substitute Equation (30) into Equation (29), then we obtain

$$\mathbf{O}(t) = \sum_{i\geq 0}\lambda_{Oi}(t)\mathbf{U}(t)\left|\Psi_{Oi}(0)\right\rangle\left\langle\Psi_{0i}(0)\right|\mathbf{U}^{+}(t) \tag{31}$$

The time derivative of the operator is

$$\begin{aligned}\dot{\mathbf{O}}(t) &= \sum_{i\geq 0}\dot{\lambda}_{Oi}(t)\mathbf{U}(t)\left|\Psi_{Oi}(0)\right\rangle\left\langle\Psi_{0i}(0)\right|\mathbf{U}^{+}(t) + \\ &+\sum_{i\geq 0}\lambda_{Oi}(t)\dot{\mathbf{U}}(t)\left|\Psi_{Oi}(0)\right\rangle\left\langle\Psi_{0i}(0)\right|\mathbf{U}^{+}(t) + \\ &+\sum_{i\geq 0}\lambda_{Oi}(t)\mathbf{U}(t)\left|\Psi_{Oi}(0)\right\rangle\left\langle\Psi_{0i}(0)\right|\dot{\mathbf{U}}^{+}(t)\end{aligned} \tag{32}$$

or by using the instantaneous eigenvectors, $|\Psi_{Oi}(t)\rangle$ can be written as

$$\begin{aligned}\dot{\mathbf{O}}(t) &= \sum_{i\geq 0}\dot{\lambda}_{Oi}(t)\mathbf{U}(t)\left|\Psi_{Oi}(0)\right\rangle\left\langle\Psi_{0i}(0)\right|\mathbf{U}^{+}(t) + \\ &+\sum_{i\geq 0}\lambda_{Oi}(t)\dot{\mathbf{U}}(t)\mathbf{U}^{+}(t)\mathbf{U}(t)\left|\Psi_{Oi}(0)\right\rangle\left\langle\Psi_{0i}(0)\right|\mathbf{U}^{+}(t) + \\ &+\sum_{i\geq 0}\lambda_{Oi}(t)\mathbf{U}(t)\left|\Psi_{Oi}(0)\right\rangle\left\langle\Psi_{0i}(0)\right|\mathbf{U}^{+}(t)\mathbf{U}(t)\dot{\mathbf{U}}^{+}(t) = \\ &= \sum_{i\geq 0}\dot{\lambda}_{Oi}(t)\left|\Psi_{Oi}(t)\right\rangle\left\langle\Psi_{0i}(t)\right| + \dot{\mathbf{U}}(t)\mathbf{U}^{+}(t)\left(\sum_{i\geq 0}\lambda_{Oi}(t)\left|\Psi_{Oi}(t)\right\rangle\left\langle\Psi_{0i}(t)\right|\right) + \\ &+\left(\sum_{i\geq 0}\lambda_{Oi}(t)\left|\Psi_{Oi}(t)\right\rangle\left\langle\Psi_{0i}(t)\right|\right)\mathbf{U}(t)\dot{\mathbf{U}}^{+}(t) = \sum_{i\geq 0}\dot{\lambda}_{Oi}(t)\left|\Psi_{Oi}(t)\right\rangle\left\langle\Psi_{0i}(t)\right| + \\ &\dot{\mathbf{U}}(t)\mathbf{U}^{+}(t)\mathbf{O}(t) + \mathbf{O}(t)\mathbf{U}(t)\dot{\mathbf{U}}^{+}(t)\end{aligned} \tag{33}$$

Using the identity $\dot{\mathbf{U}}(t)\mathbf{U}^{+}(t) = -\mathbf{U}(t)\dot{\mathbf{U}}^{+}(t)$, which is a consequence of unitarity, the general evolution equation of the Hermitian operator results

http://dx.doi.org/10.5772/61010

$$\dot{\mathbf{O}}(t)=\frac{\partial \mathbf{O}(t)}{\partial t}+\left[\dot{\mathbf{U}}(t)\mathbf{U}^{+}(t),\mathbf{O}(t)\right],$$
$$\frac{\partial \mathbf{O}(t)}{\partial t}:=\sum_{i\geq 0}\dot{\lambda}_{Oi}(t)\left|\Psi_{Oi}(t)\right\rangle\left\langle\Psi_{0i}(t)\right| \tag{34}$$

where [,] is Dirac's symbol of a quantum mechanical commutator and $\frac{\partial \mathbf{O}(t)}{\partial t}$ is the local time rate of change of the operator in the coordinate system of the instantaneous eigenvectors. This equation is universal in the meaning of its independence of the constitutive behavior of the quantum system.

3.2. The Heisenberg equation of motion of the linearly damped oscillator

The actual form of Heisenberg's dynamic equation can be constructed when the expression $\dot{\mathbf{U}}(t)\mathbf{U}^{+}(t)$ is formed from the unitary operator and the local rate of change of the operator $\frac{\partial \mathbf{O}(t)}{\partial t}$ can be connected to the constitutive properties of the studied physical system. The first term will be ordered to the unitary/reactive and the second to the dissipative dynamics. To do this, we accept Heisenberg's reinterpretation principle [14] (for the philosophical details of this principle, see [27]), which states the possibility of constructing a quantum mechanical description of a physical system whose classical description is known. In our case, the physical system is a linearly damped oscillator, for which we know its Onsagerian thermodynamic description. This description uses the Hamiltonian and the rate of energy dissipation of the system represented by the Rayleigh potential. Since the Hamiltonian is not the first integral of the motion, Bohlin's first integral could be used for the unitary dynamics. In classical quantum theory, the Hamiltonian belongs to the unitary dynamics as a constitutive property, i.e.,

$$\frac{i}{\hbar}\mathbf{H}=\dot{\mathbf{U}}(t)\mathbf{U}^{+}(t) \tag{35}$$

where the Hamilton operator **H** is a first integral of the system. In a closed physical system, the local time derivative of the observables is zero since the system is reactive. On this basis, the constitutive equations of classical quantum mechanics are

$$\frac{i}{\hbar}\mathbf{H}=\dot{\mathbf{U}}(t)\mathbf{U}^{+}(t),\quad \frac{\partial \mathbf{O}(t)}{\partial t}=0 \tag{36}$$

With these constitutive equations from Equation (34), the general evolution equation we give to Heisenberg's equation of the observable is

$$\dot{\mathbf{O}}=\frac{i}{\hbar}\left[\mathbf{H},\mathbf{O}\right] \tag{37}$$

Also, the all constitutive properties of the quantum system are contained in the Hamilton operator only, which could have originated from the Hamilton function of the classical model by means of Heisenberg's reinterpretation principle. Figure 2 shows the above-presented scheme of the deduction of Heisenberg's equation of motion.

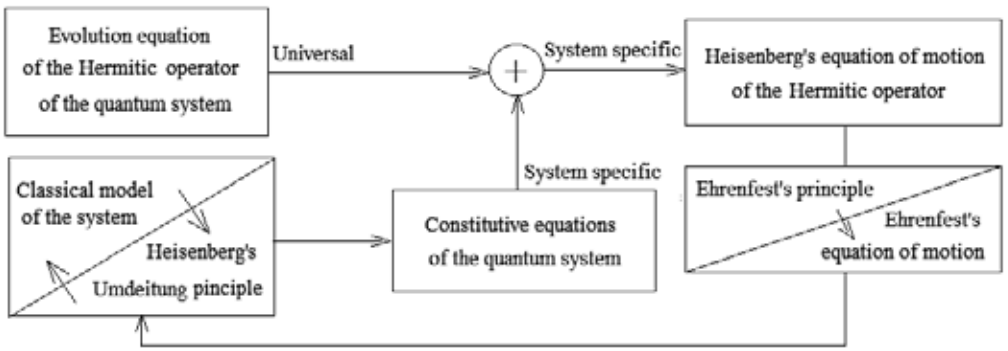

Figure 2. The schema of the deduction of Heisenberg's equation of motion of a Hermitian operator and the role of Heisenberg's reinterpretation principle.

The Hamiltonian **H** is a trivial nullifier of Dirac's commutator in this approach, so **H** is a conserved observable of motion, as was requested. To summarize, we get Heisenberg's classical Equation (37) from the general evolution Equation (34), if the $\frac{\partial \mathbf{O}(t)}{\partial t}$ local rate of change of the operator in the coordinate system of the instantaneous eigenvectors is zero. In the case of the Heisenberg's equation (Equation (37)), the entropy of the system is constant in time, as proven by von Neumann [26]. However, the entropy cannot remain constant in dissipative processes. Consequently, in a correct description of the dissipative system, it is possible to take into account the local rate of change of the operator. In the case of a damped oscillator, this means that for the time rate of change $\frac{\partial \mathbf{q}(t)}{\partial t}$, $\frac{\partial \mathbf{p}(t)}{\partial t}$ of two fundamental observables of a linearly damped oscillator must be given constitutive equations. In this way, we assume that the constitutive equations of the linearly damped oscillator are

$$\frac{\partial \mathbf{q}(t)}{\partial t} = -\beta \mathbf{q}(t), \quad \frac{\partial \mathbf{p}(t)}{\partial t} = -\beta \mathbf{p}(t), \quad \frac{i}{\hbar}\mathbf{B} = \dot{\mathbf{U}}(\mathbf{t})\mathbf{U}^{+}(\mathbf{t}) \tag{38}$$

where the Hermitian operator **B** belonging to unitary/reactive dynamics is the quantum mechanical equivalent of Bohlin's constant in Equation (26). Following Equation (26), this could be written as

$$\mathbf{B} = e^{2\beta t}\left(\frac{\mathbf{p}^2}{2m} + \frac{\beta}{2}\mathbf{pq} + \frac{\beta}{2}\mathbf{qp} + \frac{m\omega_0^2}{2}\mathbf{q}^2\right) \tag{39}$$

http://dx.doi.org/10.5772/61010

Note, the third constitutive equation of (38) is the direct consequence of Stone's theorem [28]. If we take into account these constitutive equations in the general evolution Equation (34) of the Hermitian operators, then we obtain Heisenberg's equations of motion of a quantum-damped oscillator

$$\dot{\mathbf{q}}(t) = -\beta\,\mathbf{q}(t) + \frac{i}{\hbar}\left[\mathbf{B},\mathbf{q}(t)\right], \quad \dot{\mathbf{p}}(t) = -\beta\,\mathbf{p}(t) + \frac{i}{\hbar}\left[\mathbf{B},\mathbf{p}(t)\right] \tag{40}$$

According to Heisenberg's reinterpretation principle, these equations could be interpreted by means of the Onsagerian equations of the oscillator. To do this, split the Bohlin operator (Equation (39)) into two parts. The first part contains the Hamilton operator **H** of the oscillator and the second part **D** belongs to the dissipation. We then obtain

$$\begin{aligned}
&\dot{\mathbf{q}}(t) = -\beta\,\mathbf{q}(t) + \frac{i}{\hbar}\left[\mathbf{B},\mathbf{q}(t)\right] = \left\{\frac{\partial\mathbf{q}(t)}{\partial t} + \frac{i}{\hbar}e^{2\beta t}\left[\mathbf{D},\mathbf{q}(t)\right]\right\} + \frac{i}{\hbar}e^{2\beta t}\left[\mathbf{H},\mathbf{q}(t)\right]\\
&\dot{\mathbf{p}}(t) = -\beta\,\mathbf{p}(t) + \frac{i}{\hbar}\left[\mathbf{B},\mathbf{p}(t)\right] = \left\{\frac{\partial\mathbf{p}(t)}{\partial t} + \frac{i}{\hbar}e^{2\beta t}\left[\mathbf{D},\mathbf{p}(t)\right]\right\} + \frac{i}{\hbar}e^{2\beta t}\left[\mathbf{H},\mathbf{p}(t)\right] +\\
&\mathbf{H} = \left(\frac{\mathbf{p}^2}{2m} + \frac{m\omega_0^2}{2}\mathbf{q}^2\right), \quad \mathbf{D} = \frac{\beta}{2}(\mathbf{qp} + \mathbf{pq})
\end{aligned} \tag{41}$$

The expression in {} is connected to dissipative thermodynamic current by analogy, while the currents outside the bracket are analogous to reactive currents. This interpretation, analogous to Equation (22), is supported by

$$\frac{\partial\mathbf{q}(t)}{\partial t} + \frac{i}{\hbar}e^{2\beta t}\left[\mathbf{D},\mathbf{q}(t)\right] = 0, \quad \frac{\partial\mathbf{p}(t)}{\partial t} + \frac{i}{\hbar}e^{2\beta t}\left[\mathbf{D},\mathbf{p}(t)\right] = 2\frac{\partial\mathbf{p}(t)}{\partial t} = -2\beta\,\mathbf{p}(t) \tag{42}$$

where the two first equations of Equation (38) were used. In detail, we could write

$$\begin{aligned}
&\frac{\partial\mathbf{q}}{\partial t} + \frac{i}{\hbar}e^{2\beta t}\left[\mathbf{D},\mathbf{q}\right] = -\beta\,\mathbf{q} + \frac{\beta}{2}\left(\frac{i}{\hbar}e^{2\beta t}\left[\mathbf{p},\mathbf{q}\right]\right)\mathbf{u} + \frac{\beta}{2}\mathbf{q}\left(\frac{i}{\hbar}e^{2\beta t}\left[\mathbf{p},\mathbf{q}\right]\right) = 0\\
&\frac{\partial\mathbf{p}}{\partial t} + \frac{i}{\hbar}e^{2\beta t}\left[\mathbf{D},\mathbf{p}\right] = -\beta\,\mathbf{p} - \frac{\beta}{2}\mathbf{p}\left(\frac{i}{\hbar}e^{2\beta t}\left[\mathbf{p},\mathbf{q}\right]\right) + \frac{\beta}{2}\left(\frac{i}{\hbar}e^{2\beta t}\left[\mathbf{p},\mathbf{q}\right]\right)\mathbf{p} = -2\beta\,\mathbf{p}
\end{aligned} \tag{43}$$

Now, we could see the desired interpretation analogy could be applied when the commutator relation $e^{2\beta t}[\mathbf{p},\,\mathbf{q}] = \frac{\hbar}{i}\delta$ is valid. Since every operator commutes with itself, we also have the fundamental brackets of our quantum theory

$$\{\mathbf{p},\mathbf{q}\}=\frac{\hbar}{i}\boldsymbol{\delta},\quad \{\mathbf{q},\mathbf{q}\}=0,\quad \{\mathbf{p},\mathbf{p}\}=0,\quad \{\ ,\ \}:=e^{2\beta t}[\ \ ] \tag{44}$$

Consequently, the first fundamental bracket in Equation (44) ensures that the dissipative part $\frac{\beta}{2}\mathbf{qp}+\frac{\beta}{2}\mathbf{pq}$ of the Bohlinian adds $\beta\,\mathbf{q}(t)$ term to the first equation and the $-\beta\,\mathbf{q}(t)$ term to the second equation in Equation (41). This allows to us change our attention from Bohlinian to Hamiltonian in Equation (41), obtaining

$$\begin{aligned}\dot{\mathbf{q}}(t)&=\frac{i}{\hbar}e^{2\beta t}\left[\mathbf{H}(t),\mathbf{q}(t)\right]=\frac{i}{\hbar}\left\{\mathbf{H}(t),\mathbf{q}(t)\right\},\\ \dot{\mathbf{p}}(t)&=-2\beta\,\mathbf{p}(t)+\frac{i}{\hbar}e^{2\beta t}\left[\mathbf{H}(t),\mathbf{p}(t)\right]=-2\beta\,\mathbf{p}(t)+\frac{i}{\hbar}\left\{\mathbf{H}(t),\mathbf{p}(t)\right\}\end{aligned} \tag{45}$$

which are equivalent with the equations

$$\dot{\mathbf{q}}(t)=\frac{i}{\hbar}\left\{\mathbf{H}(t),\mathbf{q}(t)\right\},\quad \dot{\mathbf{p}}(t)=\frac{i}{\hbar}\left\{\mathbf{H}(t),\mathbf{p}(t)\right\}-c\frac{i}{\hbar}\left\{\mathbf{H}(t),\mathbf{q}(t)\right\} \tag{46}$$

The quantum mechanical equations of a damped oscillator with the fundamental brackets in Equation (44), applying the rules of Lie algebra, are as follows

$$\begin{aligned}\dot{\mathbf{q}}(t)&=\frac{i}{\hbar}\left\{\mathbf{H}(t),\mathbf{q}(t)\right\}=\frac{\mathbf{p}(t)}{m},\\ \dot{\mathbf{p}}(t)&=\frac{i}{\hbar}\left\{\mathbf{H}(t),\mathbf{p}(t)\right\}-c\frac{i}{\hbar}\left\{\mathbf{H}(t),\mathbf{q}(t)\right\}=-2\beta\,\mathbf{p}(t)-m\omega_0^2\mathbf{q}(t)\end{aligned} \tag{47}$$

which are the operator differential equations version of Onsager's equations in Equation (22).

To use the Lie algebraic method in an evaluation of the above-introduced time-dependent commutators, it is assumed that the scalar time functions must necessarily be considered as ordinary numbers (for details, see [15]). In summary, according to the Onsagerian equations of the damped oscillator by application of Heisenberg's reinterpretation principle, the quantum mechanical equation of a damped oscillator in the Heisenberg picture can be obtained.

3.3. Ehrenfest theorem of a linearly damped oscillator

It is easy to show, similar to classical quantum mechanics, that the following operator analytics relations are valid [29]

$$\frac{i}{\hbar}\left\{\mathbf{H},\mathbf{q}\right\}=\frac{\partial\mathbf{H}}{\partial\mathbf{p}},\quad \frac{i}{\hbar}\left\{\mathbf{H},\mathbf{p}\right\}=-\frac{\partial\mathbf{H}}{\partial\mathbf{q}} \tag{48}$$

http://dx.doi.org/10.5772/61010

As a consequence of Equation (48), Equation (47), the quantum mechanical equations of the oscillator, could be written in the form

$$\dot{\mathbf{q}}(t)=\frac{\partial \mathbf{H}}{\partial \mathbf{p}}=\frac{\mathbf{p}(t)}{m},\quad \dot{\mathbf{p}}(t)=\frac{\partial \mathbf{H}}{\partial \mathbf{q}}-c\frac{\partial \mathbf{H}}{\partial \mathbf{p}}=-2\beta\,\mathbf{p}(t)-m\omega_0^2\mathbf{q}(t) \tag{49}$$

where the formal equivalence with Onsager's equations Equation (22) is obvious.

It is well-known in the Heisenberg picture that the expectation value of an operator is defined as

$$\langle O\rangle(t)=\langle\Psi|\mathbf{O}(t)|\Psi\rangle \tag{50}$$

where $|\Psi\rangle$ is the time-independent state vector of the oscillator. Thus, from Equation (49), the expectation values of the time rate of change of displacement and momentum can be evaluated as

$$\begin{aligned}
&\frac{d\langle q\rangle(t)}{dt}:=\langle\Psi|\dot{\mathbf{q}}|\Psi\rangle=\frac{i}{\hbar}\langle\Psi|\{\mathbf{H},\mathbf{q}\}|\Psi\rangle=\langle\Psi|\frac{\partial\mathbf{H}}{\partial\mathbf{p}}|\Psi\rangle=\frac{1}{m}\langle\Psi|\mathbf{p}|\Psi\rangle=\frac{\langle p\rangle(t)}{m},\\
&\frac{d\langle p\rangle(t)}{dt}:=\langle\Psi|\dot{\mathbf{p}}|\Psi\rangle=\frac{i}{\hbar}\langle\Psi|\{\mathbf{H},\mathbf{p}\}|\Psi\rangle-c\frac{i}{\hbar}\langle\Psi|\{\mathbf{H},\mathbf{q}\}|\Psi\rangle=\\
&=-\langle\Psi|\frac{\partial\mathbf{H}}{\partial\mathbf{q}}|\Psi\rangle-c\langle\Psi|\frac{\partial\mathbf{H}}{\partial\mathbf{p}}|\Psi\rangle=-m\omega_0^2\langle\Psi|\mathbf{q}|\Psi\rangle-\frac{c}{m}\langle\Psi|\mathbf{p}|\Psi\rangle=\\
&=-\frac{c}{m}\langle p\rangle(t)-m\omega_0^2\langle q\rangle(t)
\end{aligned} \tag{51}$$

where we take into account Equation (49). Also, the expectation values of displacement and momentum of the linearly damped oscillator obey time evolution equations, which are exactly equivalent to those of Onsager's equations (Equations (21) and (22)). This result is Ehrenfest's theorem.

4. Evaluation of the equations of the quantum linearly damped oscillator

The solutions of the operator differential equations (Equation (49)) are

$$\mathbf{q}=\sqrt{\frac{\hbar}{2m\omega_0}}\left(\mathbf{A}\,e^{-i\omega t}+\mathbf{A}^{+}e^{i\omega t}\right),\mathbf{p}=\sqrt{\frac{\hbar m}{2\omega_0}}\left(\gamma\mathbf{A}\,e^{-i\omega t}+\gamma^{*}\,\mathbf{A}^{+}e^{i\omega t}\right),\mathbf{A}=\mathbf{a}e^{-\beta t} \tag{52}$$

By substituting the above two expressions into the first fundamental commutation relation of Equation (44), the time-dependent and time-independent amplitude operators are used to obtain the following commutation relations

$$\left\{\mathbf{A},\mathbf{A}^{+}\right\}=e^{-2\beta t}\boldsymbol{\delta}\rightarrow\left[\mathbf{a},\mathbf{a}^{+}\right]=\boldsymbol{\delta} \tag{53}$$

To solve the damped oscillator problem, we have to determine the operator **A** because this should be known for the specification of displacement, momentum, and the energy of the oscillator. In the case of a nondamped oscillator, the amplitude operator can be determined from the Hamilton operator of the oscillator, which is a constant of the motion. This is, however, not true for our case; thus, we will use the Bohlin operator introduced earlier. By substituting Equation (52) into the Bohlinian Equation (39), we get

$$\begin{aligned}\mathbf{B}&=e^{2\beta t}\left(\frac{\mathbf{p}^2}{2m}+\frac{\beta}{2}\mathbf{pq}+\frac{\beta}{2}\mathbf{qp}+\frac{m\omega_0^2}{2}\mathbf{q}^2\right)=\frac{\hbar\omega_m}{2}e^{2\beta t}\left(\mathbf{AA}^{+}+\mathbf{A}^{+}\mathbf{A}\right)=\\&=\hbar\omega_m\,e^{2\beta t}\left(\mathbf{A}^{+}\mathbf{A}+\frac{1}{2e^{2\beta t}}\right)=\hbar\omega_m\left(\mathbf{a}^{+}\mathbf{a}+\frac{1}{2}\right),\quad\omega_m=\omega^2\omega_0^{-1}\end{aligned} \tag{54}$$

where the commutation relations (Equation (44)) were used. Pursuant to the above two relations, it is easy to show that the time-independent amplitude operators fulfill the equations

$$\left[\mathbf{B},\mathbf{a}^{+}\right]=\hbar w_m\mathbf{a}^{+},\left[\mathbf{a},\mathbf{B}\right]=\hbar w_m\mathbf{a} \tag{55}$$

Now, we see that if we replace the operator **B** by the Hamiltonian **H** of a simple oscillator, these equations are identical to the corresponding equations of the simple quantum oscillator [30, 31]. According to this strong analogy, we are able to determine the amplitude matrix and the matrices of the Bohlin operator **B**, the displacement operator and the momentum operator. The results are as follows:

- The operators **a**, $\mathbf{a}^{+}$, **B** and the occupation number operator $\mathbf{N}:=\mathbf{a}^{+}\mathbf{a}$ have the same eigenvectors and different eigenvalues. Since $\mathbf{N}:=\mathbf{a}^{+}\mathbf{a}$ is positive definite, **B** can possess no negative eigenvalue. The lowest eigenvalue of **B** belongs to the eigenket $|0\rangle$ of the operator **a** for which the relation $\mathbf{a}|0\rangle=0$ holds. From Equation (54), this so-called vacuum state belongs to the $\frac{1}{2}\hbar\omega_m$ zero-point Bohlinian eigenvalue and zero occupation number eigenvalue. The Bohlinian eigenvalue belonging to the eigenket $|n\rangle$ (where the occupation number is n) can be calculated in the form of $|n\rangle=\frac{\mathbf{a}^{+n}}{n!}|0\rangle$ is $(\frac{1}{2}+n)\hbar\omega_m$, while the occupation number eigenvalue is n.

http://dx.doi.org/10.5772/61010

- For the actions of the eigenket $|n\rangle$ of the ladder operators, **a** and $\mathbf{a}^+$ can be written as $\mathbf{a}|n\rangle=(n-1)|n-1\rangle$, $\mathbf{a}^+|n\rangle=(n+1)|n+1\rangle$, from which it follows for the occupation number operator that $\mathbf{N}|n\rangle=n|n\rangle$.
- The matrices of the above-introduced operators are

$$\mathbf{a}=\begin{bmatrix}0 & \sqrt{1} & 0 & .\\ 0 & 0 & \sqrt{2} & .\\ 0 & 0 & 0 & .\\ . & . & . & .\end{bmatrix}, \tag{56}$$

$$\mathbf{N}=\mathbf{a}\mathbf{a}^+=\begin{bmatrix}1 & 0 & 0 & .\\ 0 & 2 & 0 & .\\ 0 & 0 & 3 & .\\ . & . & . & .\end{bmatrix}. \tag{57}$$

on the basis of which is formed the orthonormal eigenkets

$$|0\rangle=\begin{bmatrix}1\\0\\0\\.\end{bmatrix},|1\rangle=\begin{bmatrix}0\\1\\0\\.\end{bmatrix},|2\rangle=\begin{bmatrix}0\\0\\1\\.\end{bmatrix},etc., \tag{58}$$

It is easy to see that the matrices that belong to **aa** and $\mathbf{a}^+\mathbf{a}^+$ are not diagonal. From the above equations, it follows that the rules for the time-dependent ladder and occupation number operators are

$$\begin{aligned}&\mathbf{A}^+(t)|n\rangle=e^{-\beta t}\sqrt{n+1}\,|n\rangle,\quad \mathbf{A}(t)|n\rangle=e^{-\beta t}\sqrt{n-1}\,|n-1\rangle,\\ &\mathbf{N}(t)|n\rangle=e^{-2\beta t}n|n\rangle\end{aligned} \tag{59}$$

5. Applications

5.1. Expected values of the main operators of a linearly damped oscillator

Moreover, the expected value of the occupation number in the nth energy eigenstate at time t is

$$\langle N\rangle(t) := \langle n|\mathbf{N}(t)|n\rangle = N_0 e^{-2\beta t}, \tag{60}$$

where N_0 is the occupation number at $t=0$. This result agrees well with the corresponding result derived from the system-plus-reservoir model [32]. In the energy representation, since the matrices of the operators $\mathbf{a}$, $\mathbf{a}^+$, $\mathbf{a}^2$, $\mathbf{a}^{+2}$ have zero diagonal elements, the expected values of the operators $\mathbf{A}$, $\mathbf{A}^+$, $\mathbf{A}^2$, $\mathbf{A}^{+2}$ are zero in every energy eigenstate, i.e.,

$$\langle A\rangle = 0, \quad \langle A^+\rangle = 0, \quad \langle A^2\rangle = 0, \quad \langle A^{+2}\rangle = 0 \tag{61}$$

According to these equations, the expected values $\langle q\rangle = :\langle n \mid \mathbf{q} \mid n\rangle$, $\langle p\rangle = :\langle n \mid \mathbf{p} \mid n\rangle$ of the displacement and the momentum operator

$$\mathbf{q} = \sqrt{\frac{\hbar}{2m\omega_0}}\left(\mathbf{A}e^{-i\omega t} + \mathbf{A}^+ e^{i\omega t}\right), \quad \mathbf{p} = \sqrt{\frac{\hbar m}{2\omega_0}}\left(\gamma \mathbf{A}e^{-i\omega t} + \gamma^* \mathbf{A}^+ e^{i\omega t}\right) \tag{62}$$

in the nth energy eigenstate are zero. The variance $\langle q^2\rangle = \langle n \mid \mathbf{q}^2 \mid n\rangle$, $\langle p^2\rangle = \langle n \mid \mathbf{p}^2 \mid n\rangle$ of the displacement and momentum operator in the nth energy eigenstate can be evaluated as

$$\begin{aligned}\left\langle q^2\right\rangle &= \frac{\hbar}{2m\omega_0}\left\langle \mathbf{A}\mathbf{A}^+ + \mathbf{A}^+\mathbf{A}\right\rangle = \frac{\hbar}{2m\omega_0}\left\langle \boldsymbol{\delta} e^{-2\beta t} + 2\mathbf{A}\mathbf{A}^+\right\rangle = \\ &= \frac{\hbar}{2m\omega_0} e^{-2\beta t}\left(1+2n\right)\end{aligned} \tag{63}$$

and

$$\begin{aligned}\left\langle p^2\right\rangle &= \frac{\hbar m}{2\omega_0}\left\langle \gamma\gamma^*\left(\mathbf{A}e^{i\omega t}\right)\left(\mathbf{A}^+ e^{-i\omega i}\right) + \gamma\gamma^*\left(\mathbf{A}^+ e^{-i\omega t}\right)\left(\mathbf{A}e^{i\omega t}\right)\right\rangle = \\ &= \frac{m\hbar\omega_0}{2}\left\langle \boldsymbol{\delta} e^{-2\beta t} + 2\left(\mathbf{A}e^{i\omega t}\right)\left(\mathbf{A}^+ e^{-i\omega t}\right)\right\rangle = \\ &= m\hbar\omega_0 e^{-2\beta t}\left(\frac{1}{2} + \left\langle \mathbf{a}\mathbf{a}^+\right\rangle\right) = m\hbar\omega_0 e^{-2\beta t}\left(\frac{1}{2} + n\right)\end{aligned} \tag{64}$$

where we considered the commutation relation (Equation (53)). According to these results, we obtain the expected value of the energy of the damped oscillator

$$\langle H\rangle(t) = \frac{\left\langle p^2\right\rangle}{2m} + m\omega_0^2\frac{\left\langle q^2\right\rangle}{2} = \hbar\omega_0\left\langle \boldsymbol{\delta} e^{-2\beta t} + 2\mathbf{A}\mathbf{A}^+\right\rangle = \hbar\omega_0 e^{-2\beta t}\left(n + \frac{1}{2}\right) \tag{65}$$

http://dx.doi.org/10.5772/61010

and the uncertainty relation

$$\begin{aligned}&(\Delta q)(\Delta p)=\frac{\hbar}{2}e^{-2\beta t}(2n+1),\rightarrow(\Delta q)(\Delta p)\geq\frac{\hbar}{2}e^{-2\beta t}\\&(\Delta q)^2:=\left\langle\left(\mathbf{q}-\boldsymbol{\delta}\langle q\rangle\right)^2\right\rangle=\left\langle q^2\right\rangle,\quad(\Delta p)^2:=\overline{\left(\mathbf{p}-\boldsymbol{\delta}\langle p\rangle\right)^2}=\left\langle p^2\right\rangle\end{aligned}\tag{66}$$

where we considered that $\langle q\rangle=0$, $\langle p\rangle=0$. Now, we can see Heisenberg's uncertainty relation is not fulfilled and, in the case of the simple oscillator and also when $c=0$, this relation transforms into Heisenberg's relation.

5.2. Probability description of the wave packet motion of the damped oscillator

To learn something about the time dependence of our system in a certain state $|\rangle$, we will calculate $\langle q|\rangle(t)$ and $|\langle q|\rangle|^2(t)$, which represent the probability amplitude and probability of finding the damped oscillator at q at time t in that state. In particular, it is useful to study the so-called coherent state $|a\rangle$, which is an eigenstate of the non-Hermitian time-dependent operator **A**, i.e.,

$$\mathbf{A}\,e^{i\omega t}\left|a\right\rangle=\sqrt{\frac{m\omega_0}{2\hbar}}A_0\;e^{-\beta t}e^{i\omega t}\left|a\right\rangle\tag{67}$$

We shall also calculate the probability amplitude $\Psi_a(q)=\langle q|a\rangle$ of the wave packet $|a\rangle$ at q. To do this, we shall start the following fact of bracket calculus

$$\begin{aligned}&\left\langle q\left|\mathbf{A}e^{i\omega t}\right|q\right\rangle=\int\left\langle q\left|\mathbf{A}e^{i\omega t}\right|q'\right\rangle\left\langle q'\middle|q\right\rangle dq'=\int\left\langle q\left|\mathbf{A}e^{i\omega t}\right|q'\right\rangle\Psi_a\left(q'\right)dq'=\\&=\sqrt{\frac{m\omega_0}{2\hbar}}A_0\;e^{-\beta t}e^{i\omega t}\Psi_a\left(q\right)\end{aligned}\tag{68}$$

Now, we are going to express the operator **A** by using the displacement and the momentum operator (52), so we get

$$\mathbf{A}\,e^{i\omega t}=\sqrt{\frac{2\hbar m}{\omega_0}}\,\frac{\mathbf{p}-m\gamma^*\mathbf{q}}{i\omega}\tag{69}$$

Taking this expression into (68), we obtain

$$\int\left\langle q\left|\frac{\mathbf{p}-m\gamma^*\mathbf{q}}{i}\right|q'\right\rangle\Psi_a\left(q'\right)dq'=m\omega A_0e^{-\beta t}e^{i\omega t}\Psi_a\left(q\right)\tag{70}$$

According to the following

$$\langle q|\mathbf{p}|q'\rangle = e^{-2\beta t}\frac{\hbar}{i}\frac{d\delta(q-q')}{dq} + f(q), \quad \langle q|\mathbf{q}|q'\rangle = q\delta(q-q') \tag{71}$$

coordinate representation of the operators originating from the commutation relation Equation (44) (for details, see [15]), we get

$$\left(\omega q + \frac{\hbar}{m}\frac{d}{dp}e^{-2\beta t}\right)\Psi_d(q) = \omega A_0 e^{-\beta t}e^{i\omega t}\Psi_d(q) \tag{72}$$

an ordinary differential equation, where $-cq$ is chosen for the arbitrary function $f(q)$. The solution of this equation in a normalized form is

$$\Psi_d(q) = g(t)\left(\frac{\hbar\omega}{\pi}\right)^{\frac{1}{4}} e^{-\frac{\omega m}{2\hbar}\frac{\left(q-A_0e^{-\beta t}e^{i\omega t}\right)^2}{e^{-2\beta t}}} \tag{73}$$

where the function $g(t)$ is evaluated in follows. From which the probability will be

$$\left|\Psi_a(q)\right|^2 = \left(\frac{m\omega}{2\hbar\pi}\right)^{\frac{1}{2}} e^{\beta t}e^{-\frac{m\omega}{\hbar}e^{2\beta t}\left(q-A_0e^{-\beta t}\cos(\omega t)\right)^2} \tag{74}$$

Now, we might see that this is the Gaussian distribution with the

$$\left|\Psi_a(q)\right|^2 = \frac{1}{\sqrt{2\pi}\left(\sqrt{\frac{\hbar}{m\omega}e^{-\beta t}}\right)}e^{-\frac{\left(q-A_0e^{-\beta t}\cos(\omega t)\right)^2}{2\left(\sqrt{\frac{\hbar}{m\omega}e^{-\beta t}}\right)^2}} \tag{75}$$

probability density function. Therefore, the motion of the center of the wave packet $|a\rangle$ is a damped oscillation, and its uncertainty width decreases exponentially from the initial value of $\sqrt{\frac{\hbar}{m\omega}}$ to zero (see Figure 3).

http://dx.doi.org/10.5772/61010

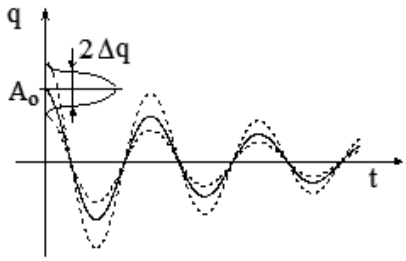

Figure 3. The evolution of the wave packet $|a\rangle$. The motion of the center of packet and its uncertainty width Δq are represented.

Now, we see that the initial uncertainty of the packet $|a\rangle$ keeps getting smaller with the progression of time and becomes negligible as $t \to \infty$. Also, the evolution of the wave packet continually proceeds toward the motion of a classic damped oscillator with the progression of time.

5.3. Calculation of wave function by matrix calculus

Resulting from Equation (52) using Equation (56), the matrix of the displacement operator in energy representation has the form

$$\langle n'|\mathbf{q}|n''\rangle := \langle n'|\mathbf{A}+\mathbf{A}^{+}|n''\rangle = \sqrt{\frac{\hbar}{2m\omega_0}} \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & \sqrt{2} & 0 & 0 \\ 0 & \sqrt{2} & 0 & \sqrt{3} & 0 \\ 0 & 0 & \sqrt{3} & 0 & \sqrt{4} \\ 0 & 0 & 0 & \sqrt{4} & 0 \\ & & & & etc. \end{bmatrix} e^{-\beta t} \tag{76}$$

Here, the displacement operator was used in a narrow sense. Next, we are going to solve the

$$\mathbf{q}|q\rangle = q|q\rangle \tag{77}$$

eigenvalue problem in terms of the $d_n = \langle n \mid q\rangle$ probability amplitudes. To do this, we rewrite the above eigenvalue equation in matrix form

$$\sum_{n'\geq 0} \langle n|\mathbf{q}|n'\rangle \langle n'|q\rangle = q\langle n|q\rangle, (n=0,1,2,\ldots) \tag{78}$$

where we take into account the $\sum_{n'\geq 0} |n'\rangle\langle n'| = \boldsymbol{\delta}$ closure relation. In explicit form, this looks like

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & \sqrt{2} & 0 & 0 \\ 0 & \sqrt{2} & 0 & \sqrt{3} & 0 \\ 0 & 0 & \sqrt{3} & 0 & \sqrt{4} \\ 0 & 0 & 0 & \sqrt{4} & 0 \\ & & & & etc. \end{bmatrix} e^{-\beta t} \begin{bmatrix} d_0 \\ d_1 \\ d_3 \\ . \\ . \\ . \end{bmatrix} = \sqrt{\frac{2m\omega_0}{\hbar}} q \begin{bmatrix} d_0 \\ d_1 \\ d_3 \\ . \\ . \\ . \end{bmatrix} \tag{79}$$

From this, we get the difference schema

$$\begin{aligned} & d_1 = \frac{q^o}{\sqrt{1}} d_0,\ d_2 = \frac{q^o}{\sqrt{2}} d_1 - \frac{1}{\sqrt{2}} d_0, \ldots, \\ & d_n = \frac{q^o}{\sqrt{n}} d_{n-1} - \frac{\sqrt{n-1}}{\sqrt{n}} d_{n-2} \\ & q^o = \sqrt{\frac{2m\omega_0}{\hbar}}\, e^{\beta t} q \end{aligned} \tag{80}$$

After some algebra, we obtain another form

$$d'_n = 2\frac{q^\circ}{\sqrt{2}} d'_{n-1} - 2(n-1) d'_{n-2}, \qquad d'_n := \sqrt{2^n n!}\, d_n \tag{81}$$

By introducing a new coordinate variable, we have the difference equation

$$d'_n = 2\hat{q} d'_{n-1} - 2(n-1) d'_{n-2}, \qquad \hat{q} := \frac{q^\circ}{\sqrt{2}} \tag{82}$$

This difference equation is satisfied by the Hermitian polynomials. Thus, we obtain

$$\langle n | q \rangle = d_n = c_0 \left(e^{\beta t} \sqrt{\frac{m\omega_0}{\hbar}} q \right) \frac{1}{\sqrt{2^n n!}} H_n \left(e^{\beta t} \sqrt{\frac{m\omega_0}{\hbar}} q \right), \tag{83}$$

where $c_0\left(e^{\beta t}\sqrt{\frac{m\omega_0}{\hbar}} q\right)$ is a function to be determined. Starting from the fact that $\langle q' \mid q'' \rangle = \delta(q' - q'')$, we get

http://dx.doi.org/10.5772/61010

$$\begin{aligned}
&\left\langle q'\middle|q''\right\rangle = \sum_n \left\langle q'\middle|n\right\rangle\left\langle n\middle|q''\right\rangle = \sum_n c_0\left(e^{\beta t}\sqrt{\frac{m\omega_0}{\hbar}}q'\right)\frac{1}{\sqrt{2^n n!}}H_n\left(e^{\beta t}\sqrt{\frac{m\omega_0}{\hbar}}q'\right)\times \\
&\times c_0^*\left(e^{\beta t}\sqrt{\frac{m\omega_0}{\hbar}}q''\right)\frac{1}{\sqrt{2^n n!}}H_n\left(e^{\beta t}\sqrt{\frac{m\omega_0}{\hbar}}q''\right) \\
&= \sum_n c_0\left(e^{\beta t}\sqrt{\frac{m\omega_0}{\hbar}}q'\right)c_0^*\left(e^{\beta t}\sqrt{\frac{m\omega_0}{\hbar}}q''\right)\frac{1}{2^n n!}H_n\left(e^{\beta t}\sqrt{\frac{m\omega_0}{\hbar}}q'\right)\times \\
&H_n\left(e^{\beta t}\sqrt{\frac{m\omega_0}{\hbar}}q''\right) = \delta\left(q'-q''\right)
\end{aligned} \tag{84}$$

By using the

$$\begin{aligned}
&\sum_n \frac{1}{2^n n!}H_n\left(e^{\beta t}\sqrt{\frac{m\omega_0}{\hbar}}q'\right)\times H_n\left(e^{\beta t}\sqrt{\frac{m\omega_0}{\hbar}}q''\right) = \\
&= \sqrt{\pi}e^{\left(e^{\beta t}\sqrt{\frac{m\omega_0}{\hbar}}q'\right)^2}\delta\left(e^{\beta t}\sqrt{\frac{m\omega_0}{\hbar}}q'-e^{\beta t}\sqrt{\frac{m\omega_0}{\hbar}}q''\right) = \\
&= \sqrt{\frac{\pi m\omega_0}{\hbar}}e^{\left(e^{\beta t}\sqrt{\frac{m\omega_0}{\hbar}}q'\right)^2}e^{\beta t}\delta\left(q''-q'\right)
\end{aligned} \tag{85}$$

relationship, the final form of Equation (83) is given by

$$\Psi_n\left(q\right) := \left\langle n\middle|q\right\rangle = d_n = e^{\frac{1}{2}\beta t}\left(\frac{m\omega_0}{\hbar\pi}\right)^{\frac{1}{4}}e^{-\frac{1}{2}e^{2\beta t}\frac{m\omega_0}{\hbar}q^2}\frac{1}{\sqrt{2^n n!}}H_n\left(e^{\beta t}\sqrt{\frac{m\omega_0}{\hbar}}q\right) \tag{86}$$

The physical meaning of the strange variable in the Hermit polynomials is that the distance of the nodes of these functions keeps getting smaller with the progression of time by the exponential law $e^{-\beta t}$. According to this result, the probability density of the nth energy state with displacement q is

$$\begin{aligned}
&\left|\Psi_n\left(q\right)\right|^2 = \left|\left\langle n\middle|q\right\rangle\right|^2 = \frac{1}{2^n n!}e^{\beta t}\left(\frac{m\omega_0}{\hbar\pi}\right)^{\frac{1}{2}}e^{-\frac{\frac{m\omega_0}{\hbar}q^2}{e^{-2\beta t}}}H_n^2\left(e^{-\beta t}\sqrt{\frac{m\omega_0}{\hbar}}q\right) \\
&= \frac{1}{\sqrt{2\pi}\left(\sqrt{\frac{\hbar}{2m\omega_0}}e^{-\beta t}\right)}e^{-\frac{q^2}{2\left(\sqrt{\frac{\hbar}{2m\omega_0}}e^{-\beta t}\right)^2}}\left[\frac{1}{2^n n!}H_n^2\left(e^{-\beta t}\sqrt{\frac{m\omega_0}{\hbar}}q\right)\right]
\end{aligned} \tag{87}$$

This result is exactly identical to the equation given by Kim and Page [33] on the basis of another theory. Now, we might see that this is the density function of a modulated Gaussian distribution, where the modulating term has finite amplitude which runs over in time, while the Gaussian distribution sharpens toward to a Dirac delta distribution. This means that the particle will get closer and closer to the equilibrium point as $t \to \infty$. From last result, we can conclude that in the case of $\beta \to 0$ we get back to the well-known wave function of the simple oscillator.

5.4. Spectrum of the energy dissipation of the linearly damped oscillator

We are going to give the frequency spectrum of radiation and explain the natural width of the spectrum line. As an atom emits photons, its energy drops and the amplitude of transition decreases over time. Therefore, the emission is not harmonic, and a spectrum occurs. We shall see that the natural width of the spectral line can be connected to the attenuation coefficient of the damped oscillator. Inversely, from the width of the spectral line, we might determine the attenuation coefficient of the oscillator.

5.4.1. Spectral density of the energy dissipation

In the first section, the time rate of energy dissipation for a damped oscillator is introduced by the Rayleigh dissipation potential. The quantum version of this quantity, i.e., the time rate of the energy dissipation operator, can be originated from Equation (14) as

$$\mathbf{R} := \frac{c}{m}\frac{1}{m}\mathbf{p}^2 = \frac{2\beta}{m}\mathbf{p}^2 \tag{88}$$

From this, it follows that the expected value of the operator of energy dissipation is

$$w_{diss} = \int_0^\infty \langle \mathbf{R} \rangle dt = \int_0^\infty \langle n|\mathbf{R}|n \rangle dt = \frac{2\beta}{m}\int_0^\infty \langle n|\mathbf{p}|^2\, n \rangle dt \tag{89}$$

Substituting the expression of the momentum operator (Equation (64)) into this equation, then the above equation has the form

$$\begin{aligned} w_{diss} &= \frac{\beta\hbar}{\omega_0}\int_0^\infty \left\langle \left(\gamma\gamma^* \left(\mathbf{A}e^{i\omega t} \right)\left(\mathbf{A}^+ e^{-i\omega t} \right) + \gamma\gamma^* \left(\mathbf{A}^+ e^{-i\omega t} \right)\left(\mathbf{A}e^{-i\omega t} \right) \right) \right\rangle dt = \\ &= \beta\hbar\omega_0 \int_0^\infty \left\langle \left(\boldsymbol{\delta} e^{-2\beta t} + 2\left(\mathbf{A}e^{-i\omega t} \right)\left(\mathbf{A}^+ e^{-i\omega t} \right) \right) \right\rangle dt = \frac{\hbar\omega_0}{2} + 2\beta\hbar\omega_0 \left\langle \mathbf{a}\mathbf{a}^+ \right\rangle \int_0^\infty e^{\gamma t} e^{\gamma^* t} dt \\ &= \frac{\hbar\omega_0}{2} + 2\beta n\hbar\omega_0 \int_0^\infty e^{\gamma t} e^{\gamma^* t} dt, \quad \gamma = \beta + i\omega \end{aligned} \tag{90}$$

http://dx.doi.org/10.5772/61010

where we assume that the occupation number is n, i.e., $n=\langle \mathbf{a}\mathbf{a}^{+}\rangle$. Evidently, the first term of this expression belongs to the vacuum fluctuation, and the second term belongs to the essential dissipative process of the damped oscillator in which the occupation number could change. We will evaluate the second term of this energy dissipation formula. According to the Parseval theorem of the Fourier transformation theory, this term of energy dissipation may be written as

$$\Delta w_{diss}\left(n\right)=2\beta n\hbar\omega_0\int_0^{\infty}e^{\gamma t}e^{\gamma^{*}t}dt=n2\beta\hbar\omega_0\int_0^{\infty}F\left[e^{\gamma t}\right]F\left[e^{\gamma t}\right]^{*}d\omega=$$
$$=n2\beta\hbar\omega_0\int_0^{\infty}\frac{d\omega'}{\left(\omega-\omega'\right)^2+\beta^2} \tag{91}$$

where $F[e^{\gamma t}]$ is the Fourier transform of $e^{\gamma t}$. Also, the spectral density of the dissipated energy is

$$\frac{2\beta\hbar\omega_0}{\left(\omega-\omega'\right)^2+\beta^2} \tag{92}$$

i.e., a Lorentz distribution about the shifted circular frequency ω of the damped oscillator. This result agrees well with corresponding result derived from the two-state atom model of Wigner and Weisskopf [34, 35] and the system-plus-reservoir model [10, 32, 36]. It is well known that the half value width $\Delta\omega$ of this distribution is

$$\Delta\omega=\beta \tag{93}$$

Now, we can see the transition from the nth occupation number state to the vacuum state, in which the oscillator will emit $n\hbar\omega_0$ energy. Indeed, from Equation (91) it follows that

$$\Delta w_{diss}\left(n\right)=n2\beta\hbar\omega_0\int_0^{\infty}\frac{d\omega'}{\left(\omega-\omega'\right)^2+\beta^2}=n\hbar\omega_0 \tag{94}$$

5.4.2. Natural width of the spectral line

The natural line width of the spectral line is a significant result of the dissipative quantum process which accompanies the spontaneous emission of an atom. We will treat this emission process in a dissipative two-state model. We consider the two states of the atom as the zeroth and the first occupation number state of a linearly damped oscillator. In this case, the spontaneous emission of a photon is the consequence of the transition from the first occupations number state to the equilibrium state of the damped oscillator. In this model, the spectrum density of the emitted photon follows from Equation (92)

$$\Delta w_{diss}\left(\omega'\right)=\frac{2\beta\hbar\omega_0}{\left(\omega-\omega'\right)^2+\beta^2} \tag{95}$$

The width of this frequency spectrum of a spontaneous emission of the atom is a direct consequence of the dissipative self-force on the atom due to the back-reaction of the emitted photon. This back-reaction of the emitted photon can be characterized by two physical quantities, namely, the frequency shift $\omega_0 \rightarrow \omega$ and the half value width $\Delta\omega = \beta$ of the spectrum. If we consider $\hbar\frac{\Delta\omega}{2}$ as the energy uncertainty ΔE of the emitted wave packet and the time constant of the emission process $\Delta t = \beta^{-1}$ as the time uncertainty, we obtain an uncertainty relation

$$\Delta E\Delta t \geq \frac{\hbar}{2\beta}\beta=\frac{\hbar}{2} \tag{96}$$

The quantum mechanical interpretation of the width of the natural spectral line should be based on this relation, in which the physical quantities ΔE and $\Delta t = \beta^{-1}$ have a precise meaning. In our model, the natural line width occurs at wavelength λ and can be calculated as

$$\Delta\lambda=\left|\Delta\frac{2\pi c}{\omega}\right|=\frac{2\pi c}{\omega^2}\Delta\omega=\frac{2\pi c}{\omega^2}\beta \tag{97}$$

where Equation (93) was used and c is the vacuum velocity of light. It is well known that in the classical dipole model of light emission, the natural line width can be calculated as

$$\Delta\lambda=\frac{4\pi\varepsilon_0}{3mc^2} \tag{98}$$

where ε_0 is the vacuum permittivity and $r_e := \frac{\varepsilon_0}{3mc^2} = 2,818\,10^{-15} m$ is the so-called classical electron radius. From the above two equations, it follows that

$$\beta=\frac{2}{3}\frac{\omega^2}{c}r_e=\frac{4\pi}{3}\frac{r_e}{\lambda}\omega \tag{99}$$

in the dipole radiation model.

http://dx.doi.org/10.5772/61010

6. Uncertainty relation of the linearly damped oscillator

The standard derivation of Heisenberg's uncertainty relation neglects the possibility that two operators **A** and **B**, say **q** and **p**,which fulfill the commutator relation

$$\left\{\mathbf{A},\mathbf{B}\right\} = i\hbar \tag{100}$$

could have a compatible component which is the first part of the trivial identity

$$\mathbf{AB} = \frac{\mathbf{AB}+\mathbf{BA}}{2} + \frac{\mathbf{AB}-\mathbf{BA}}{2} \tag{101}$$

This observation has importance when we take into account the irreversibility. Due to irreversibility, the damped oscillator proceeds to thermal equilibrium with the thermal bath. This thermal equilibrium can be characterized in terms of classical statistic theory. However, in classical statistics, random variables have a joint distribution function, which could exist in the case of quantum theory if the operators are compatible. The commutator relation (Equation (100)) is compatible this physical picture, but from Equations (100) and (101), we obtain

$$\mathbf{AB} = \frac{\mathbf{AB}+\mathbf{BA}}{2} + \boldsymbol{\delta}\frac{i\hbar}{2}e^{-2\beta t} \tag{102}$$

From this relation, in the case of $t \to \infty$, the compatibility of the operators follows, i.e.,

$$\mathbf{AB} = \frac{\mathbf{AB}+\mathbf{BA}}{2} \to \mathbf{AB} = \mathbf{BA} \tag{103}$$

In what follows, we will show that the above-mentioned arguments appear in the uncertainty relation. The variance of the Hermitian operators **A** and **B** can be calculated by the norm of the following vectors

$$f = \left(\mathbf{A} - \boldsymbol{\delta}\left\langle\mathbf{A}\right\rangle\right)\left|\Psi(0)\right\rangle,\ g = \left(\mathbf{B} - \boldsymbol{\delta}\left\langle\mathbf{B}\right\rangle\right)\left|\Psi(0)\right\rangle \tag{104}$$

Indeed, we can write

$$\begin{aligned}\left(\Delta A\right)^2 &= \left\|f\right\|^2 = \left\langle\Psi(0)\right|\left(\mathbf{A} - \boldsymbol{\delta}\left\langle A\right\rangle\right)^2\left|\Psi(0)\right\rangle \\ \left(\Delta B\right)^2 &= \left\|g\right\|^2 = \left\langle\Psi(0)\right|\left(\mathbf{B} - \boldsymbol{\delta}\left\langle B\right\rangle\right)^2\left|\Psi(0)\right\rangle\end{aligned} \tag{105}$$

where $\Psi(0)$ is the state vector of the system, $\langle A\rangle$ and $\langle B\rangle$ are the expected value of the operators defined as

$$\langle A\rangle = \langle \Psi(0)|\mathbf{A}|\Psi(0)\rangle, \langle B\rangle = \langle \Psi(0)|\mathbf{B}|\Psi(0)\rangle \tag{106}$$

Thus, the $\| f \|^2 \| g \|^2 \geq |\langle f, g\rangle|^2$ Schwarz inequality implies

$$(\Delta A)^2 (\Delta B)^2 \geq \left|\langle \Psi(0)|\left(\mathbf{A} - \boldsymbol{\delta}\langle \mathbf{A}\rangle\right)\left(\mathbf{B} - \boldsymbol{\delta}\langle \mathbf{B}\rangle\right)|\Psi(0)\rangle\right|^2 = \left|\langle AB\rangle - \langle A\rangle\langle B\rangle\right|^2 \tag{107}$$

By substituting into this expression the identity (Equation (101)), then we get

$$\begin{aligned}(\Delta A)^2 (\Delta B)^2 &\geq \left|\frac{\langle AB\rangle + \langle BA\rangle}{2} + \frac{\langle AB - BA\rangle}{2} - \langle A\rangle\langle B\rangle\right|^2 = \\ &= \left|\frac{\langle AB\rangle + \langle BA\rangle}{2} + \frac{i\hbar}{2} e^{-2\beta t} - \langle A\rangle\langle B\rangle\right|^2 = \frac{\hbar^2}{4} e^{-4\beta t} + \left(\frac{\langle AB\rangle + \langle BA\rangle}{2} - \langle A\rangle\langle B\rangle\right)^2\end{aligned} \tag{108}$$

where we take into account that the quadrate of the absolute value of a complex number is equally the sum of the quadrate of its real and imaginary parts. From the above expression, in the case of $t \to \infty$, it follows that

$$\Delta A \Delta B \geq \frac{\langle AB\rangle + \langle BA\rangle}{2} - \langle A\rangle\langle B\rangle, \quad \Delta A := +\sqrt{(\Delta A)^2}, \quad \Delta B := +\sqrt{(\Delta B)^2} \tag{109}$$

On the another hand, in this case, the commutating relation (Equation (103)) is valid; thus, we can conclude that

$$(\Delta A)(\Delta B) \geq \langle AB\rangle - \langle A\rangle\langle B\rangle \to 0 \leq \frac{\langle AB\rangle - \langle A\rangle\langle B\rangle}{(\Delta A)(\Delta B)} \leq 1 \tag{110}$$

which is the most primitive "uncertainty relation" of classical statistic theory in which the random variables have a joint distribution function. It states the simple fact that the regression coefficient is smaller than one if the random variables are not statistically independent.

In summary, we can provide a speculative interpretation of irreversibility in quantum mechanics, namely, in an irreversible quantum process. The incompatible operators proceed to compatible ones, which are submitted to the laws of classical statistic theory.

http://dx.doi.org/10.5772/61010

7. Quantum statistics of the linearly damped oscillator

It was von Neumann [26] who first dealt with the problem of the quantum statistical ensemble. The density operator is the statistical operator of a quantum statistical ensemble. In our case, the statistical ensemble is a set of linearly damped oscillators of several quantum states in contact with a heat bath with temperature T. The density operator is an operator whose eigenvalues are the classical statistical probability of the chosen microstates denoted by p_i. If the chosen microstates are denoted by $|i\rangle$, which are eigenstates of a Hermitian operator but not necessarily the eigenstate of a Bohlinian or Hamiltonian, the general density operator is written as

$$\boldsymbol{\rho} := \boldsymbol{\rho}\boldsymbol{\delta} = \sum_{i\geq 0} \boldsymbol{\rho}\left|i\right\rangle\left\langle i\right| = \sum_{i\geq 0} p_i \left|i\right\rangle\left\langle i\right|, \quad \boldsymbol{\delta} = \sum_{i\geq 0} \left|i\right\rangle\left\langle i\right| \tag{111}$$

From this definition, it follows that ρ is Hermitian and normalized

$$\boldsymbol{\rho} = \boldsymbol{\rho}^{+}, \quad Tr\boldsymbol{\rho} = 1 \tag{112}$$

In the Heisenberg picture, the density operator is time independent and is written as

$$\boldsymbol{\rho}_H = \sum_{i\geq 0} p_i(0)\left|i(0)\right\rangle\left\langle i(0)\right| \tag{113}$$

The ensemble average of an operator in the Heisenberg picture $\mathbf{A}_H(t)$ is defined as

$$\left\langle A\right\rangle(t) = Tr\left(\boldsymbol{\rho}_H \mathbf{A}_H(t)\right) = \sum_{i\geq 0} p_i(0)\left\langle i(0)\right|\mathbf{A}_H(t)\left|i(0)\right\rangle \tag{114}$$

Ensemble averages of time rate of change of the displacement and the momentum of the linearly damped oscillator can be evaluated from Equation (51) as follows

$$\begin{aligned}
&\frac{d\left\langle q\right\rangle(t)}{dt} := Tr\left(\boldsymbol{\rho}_H \dot{\mathbf{q}}\right) = \frac{i}{\hbar} Tr\left(\boldsymbol{\rho}_H \left\{\mathbf{H},\mathbf{q}\right\}\right) = Tr\left(\boldsymbol{\rho}_H \frac{\partial \mathbf{H}}{\partial \mathbf{p}}\right) = \left\langle \frac{\partial \mathbf{H}}{\partial \mathbf{p}}\right\rangle = \left\langle \frac{\mathbf{p}}{m}\right\rangle = \frac{\left\langle p\right\rangle(t)}{m} \\
&\frac{d\left\langle p\right\rangle(t)}{dt} := Tr\left(\boldsymbol{\rho}_H \dot{\mathbf{p}}\right) = \frac{i}{\hbar} Tr\left(\boldsymbol{\rho}_H \left\{\mathbf{H},\mathbf{p}\right\}\right) - c\frac{i}{\hbar} Tr\left(\boldsymbol{\rho}_H \left\{\mathbf{H},\mathbf{q}\right\}\right) = \\
&= -Tr\left(\boldsymbol{\rho}_H \frac{\partial \mathbf{H}}{\partial \mathbf{q}}\right) - cTr\left(\boldsymbol{\rho}_H \frac{\partial \mathbf{H}}{\partial \mathbf{p}}\right) = -\left\langle \frac{\partial \mathbf{H}}{\partial \mathbf{q}}\right\rangle - c\left\langle \frac{\partial \mathbf{H}}{\partial \mathbf{p}}\right\rangle = m\omega_0^2 \left\langle \mathbf{q}\right\rangle - c\left\langle \frac{\mathbf{p}}{m}\right\rangle = \\
&= -\frac{c}{m}\left\langle p\right\rangle(t) - m\omega_0^2 \left\langle q\right\rangle(t)
\end{aligned} \tag{115}$$

Here, Equation (50) was used. Now, we see that these equations are equivalent to those of the macroscopic Onsagerian equations (Equation (21) or (22)). In the Schrödinger picture, the density operator is time dependent, but the observables of the oscillator are time independent. We define this density operator as

$$\boldsymbol{\rho}_S(t) = \sum_{i\geq 0} p_i(t) \left| i(t) \right\rangle \left\langle i(t) \right| \tag{116}$$

where we allowed a time-dependent probability $p_i(t)$ to microstate $|\, i(t)\rangle$. Also in this picture, the occupation of microstates is not conserved. The ensemble average of an operator $\mathbf{A}_S$ in the Schrödinger picture is defined as

$$\langle A \rangle(t) = Tr\left(\boldsymbol{\rho}_H(t)\mathbf{A}_S\right) = \sum_{i\geq 0} p_i(t) \left\langle i(t) \right| \mathbf{A}_S \left| i(t) \right\rangle \tag{117}$$

Two ensemble averages of an observable **A** must be equal, i.e.,

$$Tr\left(\boldsymbol{\rho}_S(t)\mathbf{A}_S\right) = Tr\left(\boldsymbol{\rho}_H \mathbf{A}_H(t)\right) \tag{118}$$

from which, in the case of pure unitary dynamics, follows the well-known transformation

$$\boldsymbol{\rho}_S(t) = \mathbf{U}^+(t)\boldsymbol{\rho}_H(t)\mathbf{U}(t) \tag{119}$$

where the unitary operator $\mathbf{U}(t)$ belongs to time evolution [37]. Since, as we have seen in the case of dissipative processes, this cannot be written by unitary dynamics only, we will use this requirement in a weaker form. We require that the basic ensemble averaged equations of the damped oscillator have the same forms in each picture, i.e.,

$$\begin{aligned} \frac{d\langle q \rangle(t)}{dt} &= Tr\left(\boldsymbol{\rho}_H \dot{\mathbf{q}}\right) = Tr\left(\dot{\boldsymbol{\rho}}_S \mathbf{q}\right) = \frac{\langle p \rangle(t)}{m} \\ \frac{d\langle p \rangle(t)}{dt} &= Tr\left(\boldsymbol{\rho}_H \dot{\mathbf{p}}\right) = Tr\left(\dot{\boldsymbol{\rho}}_S \mathbf{p}\right) = -\frac{c}{m}\langle p \rangle(t) - m\omega_0^2 \langle q \rangle(t) \end{aligned} \tag{120}$$

According to this requirement, we could give the actual form of the equation of motion for the density operator in the Schrödinger picture. We will see that this equation corresponds to the Liouville–von Neumann equation in the case of dissipative processes. From Equations (113) and (119), it follows that the density operator in the Schrödinger picture could be written by a Hermitian operator in the form

http://dx.doi.org/10.5772/61010

$$\boldsymbol{\rho}_S(t) = \sum_{i\geq 0} p_i(t)\mathbf{U}^+(t)\left|i(0)\right\rangle\left\langle i(0)\right|\mathbf{U}(t) \tag{121}$$

From the general evolution equation (Equation (34)) of the Hermitian operator, the equation of motion of the density operator in Schrödinger picture could be derived as follows:

$$\begin{aligned}\dot{\boldsymbol{\rho}}_S &= \frac{d\boldsymbol{\rho}_S(t)}{dt} = \frac{\partial\boldsymbol{\rho}_S(t)}{\partial t} + \left[\dot{\mathbf{U}}^+(t)\mathbf{U}(t), \boldsymbol{\rho}_S\right] \\ \frac{\partial\boldsymbol{\rho}_S(t)}{\partial t} &:= \sum_{i\geq 0}\frac{dp_i(t)}{dt}\mathbf{U}^+(t)\left|i(0)\right\rangle\left\langle i(0)\right|\mathbf{U}(t)\end{aligned} \tag{122}$$

where in the case of a damped oscillator, the unitary transformation belongs to the Bohlin operator of Equation (38), i.e.,

$$\dot{\mathbf{U}}^+(t)\mathbf{U}(t) = -\frac{i}{\hbar}\mathbf{B} \tag{123}$$

Thus, the equation of motion of Schrödinger's density operator is

$$\dot{\boldsymbol{\rho}}_S = \frac{d\boldsymbol{\rho}_S(t)}{dt} = \frac{\partial\boldsymbol{\rho}_S(t)}{\partial t} + \frac{i}{\hbar}\left[\mathbf{B}, \boldsymbol{\rho}_S\right] = \frac{\partial\boldsymbol{\rho}_S(t)}{\partial t} + \frac{i}{\hbar}\left\{\mathbf{H}, \boldsymbol{\rho}_S\right\} + \frac{i}{\hbar}\left\{\mathbf{D}, \boldsymbol{\rho}_S\right\} \tag{124}$$

Here, similar to the Heisenberg equations (Equation (41)), we introduced the Hamiltonian and the dissipation operator, by means of the commutator $\{\ ,\ \} = e^{2\beta t}[\ ,\ ]$. To construct a constitutive equation for the local change in the density operator, one must take into account the consequences of the commutation relation (Equation (44)) and the facts (Equations (102) and (103)) by which the incompatibility of the operators show strict fading over time. A consequence of this fading property could be increasing uncertainty in the distinction of the quantum states of the oscillators. On the other hand, in a canonical ensemble, the oscillators weakly interact with each other. By this means, the occupation of the states could change in the ensemble. Thus, we can assume that the occupation of states is not conserved in the time evolution of the ensemble. As a consequence, the statistical ensemble of the oscillators could proceed to a final state in which the classical probabilities of the microstates correspond to a classical thermal equilibrium distribution. Denoted by the density operator ρ_S in the instantaneous and ρ_{Sequ} in the equilibrium final state, then the suggested linear constitutive equation is

$$\frac{\partial\boldsymbol{\rho}_S}{\partial t} = -\beta\left(\boldsymbol{\rho}_S - \boldsymbol{\rho}_{Sequ}\right), \quad Tr\left(\boldsymbol{\rho}_S\right) = Tr\left(\boldsymbol{\rho}_{Sequ}\right) \tag{125}$$

Thus, the final form of the Liouville–von Neumann Equation (124) is

$$\begin{aligned}\dot{\boldsymbol{\rho}}_S &= \frac{d\boldsymbol{\rho}_S(t)}{dt} = -\beta\left(\boldsymbol{\rho}_S - \boldsymbol{\rho}_{Sequ}\right) + \frac{i}{\hbar}\left[\mathbf{B}, \boldsymbol{\rho}_S\right] = \\ &= -\beta\left(\boldsymbol{\rho}_S - \boldsymbol{\rho}_{Sequ}\right) + \frac{i}{\hbar}\left\{\mathbf{H}, \boldsymbol{\rho}_S\right\} + \frac{i}{\hbar}\left\{\mathbf{D}, \boldsymbol{\rho}_S\right\}\end{aligned} \tag{126}$$

We will show that this evolution equation guarantees that the equivalence relation (Equation (120)) is fulfilled, the density matrix proceeds to an equilibrium state and that the entropy of the ensemble of the damped oscillator proceeds the maximum value over time, which corresponds to thermal equilibrium. Indeed, the proof of the relations in Equation (120) proceeds as follows

$$\begin{aligned}&\frac{d\langle q\rangle(t)}{dt} = Tr\left(\dot{\boldsymbol{\rho}}_S \mathbf{q}\right) = -\beta Tr\left(\left(\boldsymbol{\rho}_S - \boldsymbol{\rho}_{Sequ}\right)\mathbf{q}\right) + \frac{i}{\hbar}Tr\left(\left\{\mathbf{H}, \boldsymbol{\rho}_S\right\}\mathbf{q}\right) + \frac{i}{\hbar}Tr\left(\left\{\mathbf{D}, \boldsymbol{\rho}_S\right\}\mathbf{q}\right) = \\ &= -\beta\langle q\rangle(t) - -\beta\langle q\rangle_{equ} + \frac{i}{\hbar}Tr\left(\boldsymbol{\rho}_S\left\{\mathbf{H}, \mathbf{q}\right\}\right) + \frac{i}{\hbar}Tr\left(\boldsymbol{\rho}_S\left\{\mathbf{D}, \mathbf{q}\right\}\right) = \\ &= -\beta\langle q\rangle(t) - \beta\langle q\rangle_{equ} + Tr\left(\boldsymbol{\rho}_S\frac{\partial \mathbf{H}}{\partial \mathbf{p}}\right) + \beta Tr\left(\boldsymbol{\rho}_S\mathbf{q}\right) = -\beta\langle q\rangle_{equ} + \frac{\langle p\rangle(t)}{m} \\ &\frac{d\langle p\rangle(t)}{dt} = Tr\left(\dot{\boldsymbol{\rho}}_S \mathbf{p}\right) = -\beta Tr\left(\left(\boldsymbol{\rho}_S - \boldsymbol{\rho}_{Sequ}\right)\mathbf{p}\right) + \frac{i}{\hbar}Tr\left(\left\{\mathbf{H}, \boldsymbol{\rho}_S\right\}\mathbf{p}\right) + \frac{i}{\hbar}Tr\left(\left\{\mathbf{D}, \boldsymbol{\rho}_S\right\}\mathbf{p}\right) = \\ &= -\beta\langle p\rangle(t) - \beta\langle p\rangle_{equ} + \frac{i}{\hbar}Tr\left(\boldsymbol{\rho}_S\left\{\mathbf{H}, \mathbf{p}\right\}\right) + \frac{i}{\hbar}Tr\left(\boldsymbol{\rho}_S\left\{\mathbf{D}, \mathbf{p}\right\}\right) = \\ &= -\beta\langle p\rangle(t) - \beta\langle p\rangle_{equ} - Tr\left(\boldsymbol{\rho}_S\frac{\partial \mathbf{H}}{\partial \mathbf{q}}\right) - \beta Tr\left(\boldsymbol{\rho}_S\mathbf{p}\right) = \\ &= -\beta\langle p\rangle_{equ} - \frac{c}{m}\langle p\rangle(t) - m\omega_0^2\langle q\rangle(t)\end{aligned} \tag{127}$$

where we take into account the cyclic invariance of the trace and the facts in Equation (42).

Now we see that if we choose an ensemble of a damped oscillator in which $\langle q\rangle_{equ}=0$, $\langle q\rangle_{equ}=0$, the required equivalence of the ensemble averages is fulfilled. To prove the increase in entropy, first we introduce quantum entropy. The $k_B\rho_S\ln\rho_S$ operator is an operator whose eigenvalues are the terms of Shannon entropy $k_B\, p_i(t)\ln p_i(t)$. Thus, the Shannon entropy is the minus trace of that operator [26]

$$S(t) := -k_B\sum_{i\geq 0} p_i(t)\ln p_i(t) = -k_B Tr\left(\boldsymbol{\rho}_S\ln\boldsymbol{\rho}_S\right) \tag{128}$$

Here, k_B is the Boltzmann constant. According to this definition, the excess entropy of an ensemble, as suggested by Bedeaux and Mazur [38], is given by

http://dx.doi.org/10.5772/61010

$$S(t) - S_{equ} = -k_B Tr\left(\ln\left(\delta\boldsymbol{\rho}_S \boldsymbol{\rho}_{Sequ}^{-1}\right)\boldsymbol{\rho}_S\right), \quad \delta\boldsymbol{\rho}_S = \boldsymbol{\rho}_S - \boldsymbol{\rho}_{Sequ} \tag{129}$$

The time rate of change of the entropy in that approximation is

$$\frac{dS(t)}{dt} = -\frac{1}{2}k_B Tr\left(\left(\delta\boldsymbol{\rho}_S \boldsymbol{\rho}_{Sequ}^{-1} + \boldsymbol{\rho}_{Sequ}^{-1}\delta\boldsymbol{\rho}_S\right)\frac{d\boldsymbol{\rho}_S}{dt}\right) \tag{130}$$

Entropy production results by substituting the Liouville–von Neumann equation into this equation

$$\begin{aligned}
&\frac{dS(t)}{dt} = -\frac{1}{2}k_B Tr\left(\left(\delta\boldsymbol{\rho}_S \boldsymbol{\rho}_{Sequ}^{-1} + \boldsymbol{\rho}_{Sequ}^{-1}\delta\boldsymbol{\rho}_S\right)\frac{d\boldsymbol{\rho}_S}{dt}\right) = \\
&= -\frac{1}{2}k_B Tr\left(\left(\delta\boldsymbol{\rho}_S \boldsymbol{\rho}_{Sequ}^{-1} + \boldsymbol{\rho}_{Sequ}^{-1}\delta\boldsymbol{\rho}_S\right)\left(-\beta\,\delta\boldsymbol{\rho}_S + \frac{i}{\hbar}\{\mathbf{B}, \boldsymbol{\rho}_S\}\right)\right) = \\
&= \frac{\beta}{2}k_B Tr\left(\left(\delta\boldsymbol{\rho}_S \boldsymbol{\rho}_{Sequ}^{-1} + \boldsymbol{\rho}_{Sequ}^{-1}\delta\boldsymbol{\rho}_S\right)\delta\boldsymbol{\rho}_S\right) - \frac{1}{2}k_B Tr\left(\left(\delta\boldsymbol{\rho}_S \boldsymbol{\rho}_{Sequ}^{-1} + \boldsymbol{\rho}_{Sequ}^{-1}\delta\boldsymbol{\rho}_S\right)\left(\frac{i}{\hbar}\{\mathbf{B}, \boldsymbol{\rho}_S\}\right)\right) = \\
&= \frac{\beta}{2}k_B Tr\left(\left(\delta\boldsymbol{\rho}_S \boldsymbol{\rho}_{Sequ}^{-1} + \boldsymbol{\rho}_{Sequ}^{-1}\delta\boldsymbol{\rho}_S\right)\delta\boldsymbol{\rho}_S\right) = \beta\, k_B Tr\left(\boldsymbol{\rho}_{Sequ}^{-1}\left(\delta\boldsymbol{\rho}_S\right)^2\right) \geq 0
\end{aligned} \tag{131}$$

where the cyclic invariance of the trace and the fact that the Bohlin operator and the ρ_{Sequ}^{-1} commutator were used. The positivity of entropy production follows from the fact that the matrices of both operators ρ_{Sequ}^{-1} and $(\delta\rho_S)^2$ have nonnegative elements only. Thus, the increase in entropy is demonstrated. It can be seen from the above deduction of entropy production that pure unitary dynamics (in the case of an undamped oscillator) is isentropic and that the entropy production is a direct consequence of the unconserved property of the occupation of states. In the thermal equilibrium, from Equation (126), it follows that **B** and ρ_{Sequ} commute, i.e.,

$$\frac{d\boldsymbol{\rho}_S(t)}{dt} = 0, \quad \boldsymbol{\rho}_S = \boldsymbol{\rho}_{Sequ}, \quad \left[\mathbf{B}, \boldsymbol{\rho}_{Sequ}\right] = 0 \tag{132}$$

So the equilibrium density operator ρ_{Sequ} is a function of the Bohlinian **B**. At equilibrium, the entropy is at maximum. Now, we maximize $S = -k_B \sum_{n\geq 0} p_n \ln p_n$ under the conditions

$$Tr\left(\boldsymbol{\rho}_S\right) = \sum_{n\geq 0} p_n = 1, \ \left\langle B\right\rangle = Tr\left(\boldsymbol{\rho}_S \mathbf{B}\right) = \sum_{n\geq 0} p_n B_n = const, \ B_n = \left\langle n\middle|\mathbf{B}\middle|n\right\rangle = \hbar\omega_m\left(\frac{1}{2} + n\right) \tag{133}$$

where we use Equation (54). The necessary condition of that maximum is

$$-k_B \sum_{n\geq 0} \delta p_n (\ln p_n + 1) = 0 \tag{134}$$

where the variations δp_n are restricted by the conditions

$$\sum_{n\geq 0} \delta p_n = 0, \quad \sum_{n\geq 0} \delta p_n B_n = 0 \tag{135}$$

Applying the method of Lagrange multipliers, we get

$$\sum_{n\geq 0} \delta p_n \left[(\ln p_n + 1) + \beta B_n + \gamma \right] = 0 \tag{136}$$

So

$$p_n = e^{-\beta B_n - \gamma - 1} \tag{137}$$

From the first equation of the conditions (133), the normalized version of the probability distribution is obtained

$$p_n = \frac{e^{-\beta B_n}}{\sum_{i\geq 0} e^{-\beta B_i}} \tag{138}$$

Choosing $\beta = \frac{1}{k_B T}$ as usual, then we get Gibbs' canonical distribution for the occupation probabilities

$$p_n = \frac{e^{-\beta B_n}}{\sum_{i\geq 0} e^{-\beta B_i}} \tag{139}$$

Introducing the partition function by definition

$$Z := Tr\left(e^{-\beta \mathbf{B}}\right) = \sum_{i\geq 0} e^{-\beta B_i} \tag{140}$$

http://dx.doi.org/10.5772/61010

then we get the equilibrium density operator

$$\boldsymbol{\rho}_{Sequ} = \frac{e^{-\beta \mathbf{B}}}{Z} \tag{141}$$

Thus, the equilibrium ensemble average of an operator $\mathbf{A}_S$ can be written as

$$\langle A \rangle = Tr\left(\boldsymbol{\rho}_{Sequ}\mathbf{A}\right) = \frac{Tr\left(e^{-\beta \mathbf{B}}\mathbf{A}\right)}{Z} = \frac{\sum_{n\geq 0}\langle n|\mathbf{A}|n\rangle e^{-\beta E_n}}{\sum_{i\geq 0} e^{-\beta B_i}} \tag{142}$$

In particular, for the ensemble average of the Bohlinian, this is

$$\langle B \rangle = \frac{\sum_{n\geq 0} B_n e^{-\beta E_n}}{\sum_{n\geq 0} e^{-\beta B_n}} = -\frac{\partial \ln Z}{\partial \beta} = \hbar\omega_m \left(\frac{1}{2} + \frac{e^{-\frac{\hbar\omega_m}{k_B T}}}{1 - e^{-\frac{\hbar\omega_m}{k_B T}}} \right) = \frac{\hbar\omega_m}{2} \coth \frac{\hbar\omega_m}{k_B T} \tag{143}$$

Introducing the free energy by definition $F = -k_B TZ$, then in terms of free energy, the Gibbs distribution is written as usual

$$p_n = e^{\frac{F - B_n}{k_B T}} \tag{144}$$

Substituting this into the definition equation of entropy (128), then we get

$$S = \frac{\langle B \rangle - F}{T} \rightarrow F = \langle B \rangle - TS \tag{145}$$

Thus, the ensemble average of the Bohlinian is the equilibrium internal energy. It is evident that the actual choice of the angular frequency ω_m in the Bohlinian is a convention. It depends on the normalization of Bohlin's constant (Equation (26)). What is the correct angular frequency? It seems from the physical aspect that the correct choice is that the angular frequency is ω_0. In this case, Bohlin's constant of motion corresponds to the maximum free energy of the linearly damped oscillator measured at time $t = 0$, so it is the exergy of the linearly damped oscillator.

8. Wave equation of the linearly damped oscillator

From the above-presented theory, we can conclude that an ensemble from a pure state always proceeds to a mixed state a consequence of irreversibility. Thus, it is impossible to describe the evolution of the pure state of a damped oscillator in the Schrödinger picture. Consequently, it is impossible to construct a linear Schrödinger equation in which the position and the momentum operator are time independent.

However, when the operators are time dependent, the model could show similarities to Schrodinger's interpretation, which we show below.

In the case of s linearly damped oscillator, the transformation of the Heisenberg picture into the Schrödinger picture by the method applied in classical quantum theory is impossible because the operator has a time-dependent part due to the dissipative process. Thus, a new way must be found to construct the wave equation of the oscillator. Kostin introduced a supplementary dissipation potential into his wave equation and constructed this dissipation potential by an assumption that the energy eigenvalues of the oscillator decay exponentially over time [39]. In Kostin's version of the wave equation, the operators are time independent, but the dissipation potential is nonlinear with respect to the wave function. In our theory, it is assumed that the abstract wave equation of the linearly damped oscillator has the form

$$i\hbar\frac{d\left|\Psi\left(t\right)\right\rangle}{dt}=\left(\mathbf{H}-\mathbf{D}\right)\left|\Psi\right\rangle \tag{146}$$

where the Hamiltonian **H** and the dissipative term D has the same mathematical form as in the Bohlinian, i.e.,

$$\mathbf{H}\left(\hat{\mathbf{p}},\hat{\mathbf{q}}\right)=\frac{\hat{\mathbf{p}}^{2}}{2m}+\frac{m\omega_{0}^{2}}{2}\hat{\mathbf{q}}^{2},\quad \mathbf{D}\left(\hat{\mathbf{p}},\hat{\mathbf{q}}\right)=\frac{\beta}{2}\left(\hat{\mathbf{p}}\hat{\mathbf{q}}+\hat{\mathbf{q}}\hat{\mathbf{q}}\right), \tag{147}$$

The operators $\hat{\mathbf{p}}$, $\hat{\mathbf{q}}$ in this picture are time dependent and satisfy the classical fundamental commutators

$$\left[\hat{\mathbf{p}},\hat{\mathbf{q}}\right]=\frac{\hbar}{i}\boldsymbol{\delta},\left[\hat{\mathbf{p}},\hat{\mathbf{p}}\right]=\left[\hat{\mathbf{q}},\hat{\mathbf{q}}\right]=\mathbf{0} \tag{148}$$

The time derivative in Equation (146) is "material" (in the sense of continuum mechanics) because of the time dependence of the observable $\hat{\mathbf{q}}$. To construct a wave equation, first we rewrite this abstract wave equation in the eigenbase $|q\rangle$ of the position operator. To do this, consider the following one-dimensional eigenvalue problem

http://dx.doi.org/10.5772/61010

$$\hat{\mathbf{q}}\left|q'\right\rangle = q\left|q'\right\rangle \tag{149}$$

which could constitute a continuous spectrum. Thus, we must write the orthonormality condition and completeness relation for the eigenvectors as follows

$$\left\langle q\middle|q'\right\rangle = \delta\left(q-q'\right), \quad \int\left|q'\right\rangle\left\langle q'\right|dq' = \boldsymbol{\delta} \tag{150}$$

where $\delta(q-q')$ is Dirac's distribution and $\boldsymbol{\delta}$ is the unity operator. The wave function is the probability amplitude that a position measurement on the damped oscillator in state $|\Psi(t)\rangle$ will yield an eigenvalue q, mathematically

$$\Psi\left(q,t\right) = \left\langle q\middle|\Psi\left(t\right)\right\rangle \tag{151}$$

Inserting Equation (150) of the unity operator in the abstract wave Equation (146) and projecting from the left with $\langle q|$, then we obtain

$$\begin{aligned} &i\hbar\left\langle q\right|\frac{a\left|\Psi\left(t\right)\right\rangle}{dt} = \frac{d\left\langle q\middle|\Psi\left(t\right)\right\rangle}{dt} = \frac{d\Psi\left(q,t\right)}{dt} = \left\langle q\middle|\mathbf{H}\left(\hat{\mathbf{p}},\hat{\mathbf{q}}\right)\right.\int\left|q'\right\rangle\left\langle q'\middle|\Psi\left(t\right)\right\rangle dq' + \\ &-\left\langle q\middle|\mathbf{D}\left(\hat{\mathbf{p}},\hat{\mathbf{q}}\right)\right.\int\left|q'\right\rangle\left\langle q'\middle|\Psi\left(t\right)\right\rangle dq' \end{aligned} \tag{152}$$

We chose the differential operator representation for the time-dependent operators in the eigenbase $|q\rangle$ of the position operator in the form

$$\hat{\mathbf{p}} = \frac{\hbar}{i}\frac{\partial}{\partial e^{\beta t}q}, \quad \hat{\mathbf{q}} = qe^{\beta t} \tag{153}$$

which resulted in the following wave equation

$$\begin{aligned} &i\hbar\frac{d\Psi\left(e^{\beta t}q,t\right)}{dt} = \left(H-D\right)\Psi\left(e^{\beta t}q,t\right), \\ &H = -\frac{\hbar^2}{2m}\frac{d^2}{d\left(e^{\beta t}q\right)^2} + \frac{m\omega_0^2}{2}\left(e^{\beta t}q\right)^2, \quad D = -i\,\hbar\left(\frac{\beta}{2} + \beta\left(e^{\beta t}q\right)\frac{d}{d\left(e^{\beta t}q\right)}\right) \end{aligned} \tag{154}$$

which is a linear partial differential equation. To construct the eigenvalue problem that belongs to this wave equation, we chose the wave function as

$$\Psi_n = e^{(\frac{E_n}{i\hbar}+\frac{\beta}{2})t}\Phi_n\left(e^{\beta t}q\right) \tag{155}$$

With this wave function, the eigenvalue equation

$$-\frac{\hbar^2}{2m}\frac{d^2\Phi_n}{d\left(e^{\beta t}q\right)^2}+\frac{m\omega_0^2}{2}\left(e^{\beta t}q\right)^2\Phi_n = E_n\Phi_n \tag{156}$$

is obtained from the wave equation because the equation

$$\begin{aligned} -D\left(e^{\frac{\beta}{2}t}\Phi_n\right) &= i\hbar\beta\left(\frac{1}{2}+\left(qe^{\beta t}\right)\frac{d}{d\left(qe^{\beta t}\right)}\right)\left(e^{\frac{\beta}{2}t}\Phi_n\right) = i\hbar\frac{de^{\frac{\beta}{2}t}\Phi_n}{dt} = \\ &= i\hbar\frac{\beta}{2}e^{\frac{\beta}{2}t}\Phi_n + i\hbar e^{\frac{\beta}{2}t}\frac{\partial\Phi_n}{\partial\left(qe^{\beta t}\right)}q\frac{de^{\beta t}}{dt} \end{aligned} \tag{157}$$

is satisfied identically for every eigenfunction Φ_n. By introducing a new variable into the eigenvalue Equation (156) defined as $\xi := \sqrt{\frac{m\omega_0}{\hbar}}e^{\beta t}q$, then we get

$$\frac{d^2\Phi_n}{d\xi^2}+\left(\frac{2E_n}{\hbar\omega_0}-\xi^2\right)\Phi_n = 0 \tag{158}$$

We chose the eigenfunction Φ_n in the form

$$\Phi_n = e^{-\frac{\xi^2}{2}}\psi_n \tag{159}$$

then we obtained the differential equation

$$\frac{d^2\psi_n}{d\xi^2}-2\xi\psi_n+\left(\frac{2E_n}{\hbar\omega_0}-1\right)\psi_n = 0 \tag{160}$$

http://dx.doi.org/10.5772/61010

which has a solution in terms of Hermitian polynoms if the

$$\frac{2E_n}{\hbar\omega_0} - 1 = 2n \rightarrow E_n = \left(n + \frac{1}{2}\right)\hbar\omega_0, \quad n = 1,2,3... \tag{161}$$

relation is fulfilled. With these, the solution of the wave equation is obtained as follows:

$$\Psi_n(\xi,t) = e^{(\frac{E_n}{i\hbar} + \frac{\beta}{2})t}\Phi_n\left(e^{\beta t}q\right) \propto e^{-i\left(n+\frac{1}{2}\right)\omega_0}e^{\frac{\beta}{2}t}e^{-\frac{\xi^2}{2}}H_n(\xi), \quad \xi := \sqrt{\frac{m\omega_0}{\hbar}}e^{\beta t}q \tag{162}$$

from which the probability density function of the oscillator has the form

$$\left|\Psi_n(q)\right|^2 = \frac{1}{2^n n!}\frac{1}{\sqrt{2\pi}\left(\sqrt{\frac{\hbar}{2m\omega_0}}e^{-\beta t}\right)}e^{-\frac{q^2}{2\left(\sqrt{\frac{\hbar}{2m\omega_0}}e^{-\beta t}\right)^2}}\left[H_n{}^2\left(e^{\beta t}\sqrt{\frac{m\omega_0}{\hbar}}q\right)\right] \tag{163}$$

which is exactly identical to Equation (87) resulting from the Heisenberg picture of the damped oscillator and the equation given by Kim and Page [33] using another theory. Due to this correspondence, the quantum decoherence of linearly damped oscillators could be described in the same way as done in the publication by Kim et al. [40].

Author details

Gyula Vincze* and Andras Szasz*

*Address all correspondence to: biotech@gek.szie.hu

Department of Biotechnics, St. Istvan University, Godollo, Hungary

References

[1] Bateman, H. (1931). On dissipative systems and related variational principles. Phys. Rev. 38:815.

[2] Feynman, R.P., Vernon, F.L. (1963). The theory of a general quantum system interacting with a linear dissipative system. Ann. Phys. 24:118–173.

[3] Weiss, U. (2008). Quantum Dissipative Systems, 3rd edition. Singapore: World Scientific.

[4] Ingold, G.L. (2012). Thermodynamic anomaly of the free damped quantum particle: the bath perspective. Eur. Phys. J. B. 85:30.

[5] Philbin, T.G. (2012). Quantum dynamics of the damped harmonic oscillator. New J. Phys. 14:083043.

[6] Chung-In, Um. et al. (2002). The quantum damped harmonic oscillator. Physics Report.362:63–192.

[7] Caldirola, P. (1941). Forze non conservative nellameccanicaquantistica. NuovoCimento. 18:393.

[8] Kanai, E. (1948). On the quantization of the dissipative systems.Prog. Theor. Phys. 3: 440.

[9] Cadirola, P. (1983). Quantum theory of nonconservative systems. NuovoCimento B. 77(2):241–262.

[10] Ullersma, P. (1966). An exactly solvable model for Brownian motion. Physica. 23, 27, 56, 74, 90.

[11] Caldeira, A. O., Leggett, A. J. (1981). Influence of dissipation on quantum tunneling in macroscopic systems. Phys. Rev. Lett. 46:211–214.

[12] Ford, G. W., O'Connell R. F. (1996). There is no quantum regression theorem. Phys. Rev. Lett. 77(5):798–801.

[13] Onsager, L. (1931). Reciprocal relations in irreversible processes. II. Phys. Rev. 38:2265.

[14] Heisenberg, W. (1925). Über die quantentheoretischeUmdeutungkinematischer und mechanischerBeziehungen. Z. Phys. 33:879–893.

[15] Vincze, Gy., Szász, A. (2014). Rosen–Chambers variationaltheory of linearly-damped oscillator. J. Adv. Phys. 4(1):404.

[16] Rosen, P. (1954). Use of restricted variational principles for the solution of differential equations. J. Appl. Phys. 25(3):336.

[17] Rosen, P. (1953). On variational principles for irreversible processes. J. Chem. Phys. 21:1220–1221.

[18] Chambers, L. (1956). A variational principle for the conduction of heat. Quart. J. Mech. Appl. Math.9:234.

[19] Meixner, J. (1963). Thermodynamics of electrical networks and the Onsager–Casmir reciprocal relations. J. Math. Phys. 4:154.

[20] Meixner, J. (1966). Network theory and its relation to thermodynamics. In: Proceedings of the Symposium on Generalized Networks. Polytechnic Press of the Polytechnic Institute of Brooklyn, New York. pp.13–25.

[21] Perelson, A. (1975). Network Thermodynamics an Overview. Biophys. J. Jul 1975; 15(7):667–685.

[22] Onsager, L. (1931). Reciprocal relations in irreversible processes. II. Phys. Rev. 38:2265.

[23] Callen, H. and Welton, T. (1951). Irreversibility and generalized noise. Phys. Rev. 83:34–40.

[24] Kubo, R. (1966). The fluctuation-dissipation theorem. Rep. Prog. Phys. 29:225.

[25] Bohlin K. (1911). Note sur le problème des deux corps et sur une intégration nouvelle dans le problème des trois corps, Bull. Astr. 28:144.

[26] von Neumann, J. (1932).MathematischeGrundlagen der Quantenmechanik. Berlin: Springer, 1932; von Neumann, J. (1955). Mathematical Foundations of Quantum Mechanics. Princeton University Press.

[27] Landsman, N.P. (2005). Between classical and quantum. arXiv: quant-ph/0506082v2, 25 Jul.

[28] Stone, M. H. (1930). Linear transformations in Hilbert space. III. operational methods and group theory. ProcNatl AcadSciUSA. 16:172–175.

[29] Pauli, W. (1958). Die AllgemeinePrinzipien der Wellenmechanik. In: Handbuch der Physik, Band 5.,Teil 1. (pp. 1–168). Berlin: Springer-Verlag; Pauli, W. (1980). General principles of quantum mechanics (translated by P. Achuthan and K. Venkatesan). Berlin: Springer Verlag(1980).

[30] Liboff, R. (2002). Introductory Quantum Mechanics, 4th edition. Addison-Wesley.

[31] Griffiths, D. (2004). Introduction to Quantum Mechanics, 2nd ed.Prentice Hall.

[32] Van Kampen, N.G. (2004). Stochastic Processes in Physics and Chemistry.Elsevier.

[33] Kim, S.P., Page, D.N. (2001). Classical and quantum action-phase variables for time-dependent oscillators. Phys. Rev. A. 64:012104.

[34] Weisskopf, V., Wigner, E. (1930). Berechnungen der NatürlichenLinienbreite auf Grund der DiracschenLichttheorie. Z. Phys.63:54–73.

[35] Heitler, W. (1954). The Quantum Theory of Radiations. Oxford: Clarendon Press.

[36] Dekker, H. (1981). Classical and quantum mechanics of the damped harmonic oscillator. Physics Report. 80:1.

[37] Merzbacher, E. (1980). Quantum Mechanics. Wiley International.

[38] Bedeaux, D., Mazur, P. (2001). Mesoscopic non-equilibrium thermodynamics for quantum systems. Physica A. 298:81–100.

[39] Kostin, M.D. (1972). On the Schrödinger–Langevin equation. J. Chem. Phys. 57:3589.

[40] Kim, S.P. et al. (2002). Decoherence of Quantum Damped Oscillators. arXiv:quant-ph/0202089v1 15 Feb.

Chapter 4

Linear Approximation of Efficiency for Similar Non-Endoreversible Cycles to the Carnot Cycle

Delfino Ladino-Luna, Ricardo T. Páez-Hernández and Pedro Portillo-Díaz

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/61011

Abstract

In the present paper the non-endoreversible Curzon-Ahlborn, Stirling and Ericsson cycles as models of thermal engines are discussed from the viewpoint of finite time thermodynamics. That is, it is propose the existence of a finite time of heat transfer for isothermal processes, but the cycles are analyzed assuming they are not endoreversible cycles, through a factor that represents the internal ireversibilities of them, so that the proposed heat engine models have efficiency closer to real engines. Some results of previous papers are used, and from the get expressions for the power output function and ecological function a methodology to obtain a linear approximation of efficiency including adequate parameters are shown, similar to those obtained in that previous paper used. Variable changes are made right, like those used previously.

Keywords: finite time thermodynamics, power output, ecological function, efficiency

1. Introduction

A valuable tool for validating and improving knowledge of nature is using models. A scientific model is an abstract, conceptual, graphic, or visual representation of phenomena, systems, or processes to analyze, describe, explain, simulate, explore, control, and predict these phenomena or processes. A model allows determining a final result from appropriate data. The creation of models is essential for all scientific activity. Moreover, a given physics theory is a model for studying the behavior of a complete system. The model is applied in all areas of physics, reducing the observed behavior to more basic fundamental facts, and helps to explain and predict the behavior of physical systems under different circumstances.

In classical equilibrium thermodynamics, the simplest model of an engine that converts heat into work is the Carnot cycle. The behavior of a heat engine working between two heat

reservoirs, modeled as this cycle, is expressed by the relation between the efficiency η and the ratio of temperatures of the heat reservoirs, T_C/T_H, with $0<T_C/T_H<1$, the Carnot efficiency $\eta_C=1-T_C/T_H$. The temperatures of reservoirs, cold and hot, respectively, are T_C and T_H (equal to those of heat engine), and η_C is a physical limit for any heat engine.

A more realistic cycle than the Carnot cycle is a modified cycle taking into account the processes time of heat transfer between the system and its surroundings, in which the working temperatures are different of those its reservoirs [1], obtaining the efficiency $\eta_{CAN}=1-\sqrt{T_C/T_H}$, first found in references [2] and [3], and known as Curzon–Ahlborn–Novikov–Chambadal efficiency. At present, the duration of heat transfer processes is important. Based on this model, at the end of the last century, a theory was developed as an extension of classical equilibrium thermodynamics, the finite time thermodynamics, in which the duration of the exchange processes heat becomes important.

Two operating regimens of a heat engine with the same type of parameters have been established: maximum power output regimen as in [1] and maximum effective power regimen, taking into account the entropy production through a function called ecological function, which represents the relationship between power output P and entropy production σ advanced in [4]. It is worth noting that there are other operating regimens such as maximum cycle efficiency or minimum entropy production. Thus, power output has been maximized in [1,5-7] among others, entropy production has been minimized in [8-10] among others, and the so-called ecological function has been maximized in [4, 11-13] among others. Also, cycles including internal irreversibilities in various aspects of operation of thermal engines have been analyzed in [14-19] and others, and the regions of existence of the objective functions listed above have been analyzed by a limited number of publications, in [9,20,21] and others. Notice that in almost all the above references, the time of the adiabatic processes in the so-called Curzon–Ahlborn cycle is assumed irrelevant because this time is considered very small compared with the total time of cycle. Nevertheless, a meticulous examination on the behavior of real engines leads to take into account the time of these adiabatic processes because these processes are only an approach to real processes in which there is no heat transfer.

An alternative to analyze the Curzon–Ahlborn cycle, taking into account some effects that are nonideal to the adiabatic processes through the time of these processes, is the model proposed in [5] and in [7]. It allows to find the efficiency of a cycle as a function of the compression ratio, $r_C=V_{\max}/V_{\min}$. When $r_C\to\infty$, $V_{\max}>>V_{\min}$, the Curzon–Ahlborn–Novikov–Chambadal efficiency is recovered. The non-endoreversible Curzon and Ahlborn cycle can be analyzed by means of the so-called non-endoreversibility parameter I_S, defined first in [14] and later in [15] and in [16], which can be used to analyze diverse particularities of cycles. Furthermore, this parameter leads to equality instead of Clausius inequality [14].

In the present paper, the performance of a non-endoreversible heat engine modeled as a Curzon–Ahlborn cycle is analyzed. The procedure in [5] is combined with the procedure in [16], arriving to linear approaches of the efficiency as a function of a parameter that contains the compression ratio in both regimens maximum power output and maximum ecological function. From the limit values of the non-endoreversibility parameter and the compression

http://dx.doi.org/10.5772/61011

ratio, the known expressions of the efficiency found in the literature of finite-time thermodynamics are recovered. Also, an analysis of the Stirling and Ericsson cycles is made, when the existence of a finite time for the heat transfer for isothermal processes is assumed, and assuming they are not endoreversible cycles, through the non-endoreversibility parameter that represents internal irreversibilities of them. Some results in [22] are used, and from the expressions obtained for the power output function and ecological function, the methodology to obtain a linear approximation of efficiency including an adequate parameter is shown, similar to those used in case of the Curzon–Ahlborn cycle. Variable changes are made right, like those used in [5] and in [23,24]. In order to make the present paper self-contained, a review of results for instantaneous adiabatic case is presented. All quantities have been taken in the International System of Measurement.

2. Linear approximation of efficiency: endoreversible Curzon–Ahlborn cycle

In a previous published chapter by InTech [25], we devoted to analyze the Curzon and Ahlborn cycle under the following conditions: without internal irreversibilities and non instantaneous adiabats. We have shown some results in case of the Newton heat transfer law (Newton cooling law) and the Dulong and Petit heat transfer law, namely, heat transfer law like $dQ/dt \propto (\Delta T)^k$, $k=5/4$. Hence, we begin with a summary of the cited chapter.

2.1. Known results and basic assumptions

Since the pioneer paper [1], the so-called finite time thermodynamics has been development. They proposed a model of thermal engine shown in Figure 1, which has the mentioned Curzon–Ahlborn–Novikov–Chambadal efficiency, as a function of the cold reservoir temperature T_C and the hot reservoir temperature T_H, as follows:

$$\eta_{CAN} = 1 - \sqrt{T_C / T_H}, \tag{1}$$

In this cycle, $Q_H / T_{HW} = Q_C / T_{CW}$ is fulfilled. The entropy production during the exchange of heat between the system and its reservoirs is only taken into account. The working temperatures of substance are T_{HW} and T_{CW}, being $T_C < T_{CW} < T_{HW} < T_H$. In contrast, the Carnot efficiency is obtained when the temperatures of reservoirs are the same as the temperatures of the engine, which means $T_{HW} = T_H$ and $T_C = T_{CW}$ in Figure 1, namely,

$$\eta_C = 1 - \frac{T_C}{T_H} = 1 - \frac{T_{CW}}{T_{CH}} \tag{2}$$

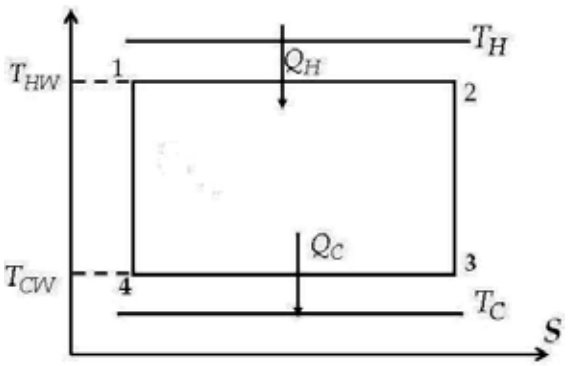

Figure 1. Curzon and Ahlborn cycle in the entropy S vs temperature T plane.

Equation (1) has been obtained at maximum power output regimen and recovered later by some procedures [5,10,26,27] among others. Moreover, in [4] was advanced an optimization criterion of merit for the Curzon and Ahlborn cycle, taking into account the entropy production, the ecological criterion, by maximization of the ecological function,

$$E = P - T_C\sigma, \tag{3}$$

where P is the power output, T_C is the temperature of cold reservoir, and σ is the total entropy production. The efficiency of Curzon and Ahlborn cycle now can be written as

$$\eta_E = 1 - \sqrt{\left(\varepsilon^2 + \varepsilon\right)/2}. \tag{4}$$

By contrast, following the procedure in [5], the form of the ecological function and its efficiency was found using the Newton heat transfer law and ideal gas as working substance in [12] and using the Dulong–Petit heat transfer law for ideal gas as working substance in [28]. Hence, as the upper limit of the efficiency of any heat engine is the Carnot efficiency, the temperatures of the reservoir equal those of the heat engine. Thus, the definition of efficiency of an engine working in cycles leads to the Carnot efficiency, fulfilling

$$\frac{Q_H - Q_C}{Q_H} \le 1 - \frac{T_{CW}}{T_{HW}}. \tag{5}$$

With $\varepsilon \equiv T_C / T_H$, the following equations can be written: Carnot efficiency, $\eta_C \equiv 1 - \varepsilon$; Curzon–Ahlborn–Novikov–Chambadal efficiency: $\eta_{CAN} = 1 - \sqrt{\varepsilon}$; and ecological efficiency: $\eta_E = 1 - \sqrt{(\varepsilon^2 + \varepsilon)/2}$. Any efficiency can be written as

http://dx.doi.org/10.5772/61011

$$\eta = 1 - z(\varepsilon). \tag{6}$$

Thus, the problem of finding the efficiency of a heat engine modeled as a Curzon–Ahlborn cycle, maximizing power output or maximizing ecological function, becomes the problem of finding a function $z = z(\varepsilon)$. Substituting $z = z(\varepsilon)$ in Equation (6), one has

$$\eta = \eta(\varepsilon) \tag{7}$$

Similar results are obtained with a nonlinear heat transfer, like the Dulong and Petit heat transfer. Assuming the same thermal conductance α in two isothermal processes of the Curzon–Ahlborn cycle, the heat exchanged between the engine and its reservoirs could be in general as

$$\frac{dQ_H}{dt} = \alpha\left(T_H - T_{HM}\right)^k \quad \text{and} \quad \frac{dQ_H}{dt} = \alpha\left(T_C - T_{CW}\right)^k \qquad , k \geq 1. \tag{8}$$

By contrast, assuming the heat flows Q_H and Q_C, given by Newton's heat transfer law, the case $k = 1$ in Equation (8), the power output becomes

$$P = \frac{\alpha T_H\left(1 - z\right)\left[1 + \lambda \ln z\right]}{\frac{1}{1-u} + \frac{z}{uz - \varepsilon}}, \tag{9}$$

where R is the general constant of gases. The parameter $\gamma \equiv C_P / C_V$ has been used, and also the variables $u = T_{HW} / T_H$ and $z = T_{CW} / T_{HW}$ from which we obtain $P = P(u, z)$. The adiabatic processes are noninstantaneous. In fact, the total time of cycle is

$$t_{TOT} = t_1 + t_2 + t_3 + t_4, \tag{10}$$

being the times for the isothermal processes,

$$t_1 = \frac{RT_{HW}}{\alpha\left(T_H - T_{HW}\right)} \ln \frac{V_2}{V_1} \text{ and } t_3 = \frac{RT_{CW}}{\alpha\left(T_{CW} - T_C\right)} \ln \frac{V_4}{V_3}, \tag{11}$$

and the times for the adiabatic processes have been assumed to be

$$t_2 = f_1 \ln \frac{V_3}{V_2} \text{ and } t_4 = f_2 \ln \frac{V_4}{V_1}, \tag{12}$$

with

$$f_1 \equiv \frac{RT_{HW}}{\alpha\left(T_H - T_{HW}\right)} \text{ and } f_2 \equiv \frac{RT_{CW}}{\alpha\left(T_{CWw} - T_C\right)}. \tag{13}$$

The maximization conditions $\partial P/\partial u = 0$ and $\partial P/\partial z = 0$ lead to obtain (or, permit obtain)

$$u = \frac{z+\varepsilon}{2z} \text{ and } \left(z^2 - \varepsilon\right)\left(1 + \lambda \ln z\right) = \lambda\left(z - \varepsilon\right)\left(1 - z\right), \tag{14}$$

where λ represents the external parameter $\lambda = [(\gamma - 1)\ln(V_3/V_1)]^{-1}$, meaning that

$$P_{\max} = P_{\max}\left(u(z), z\right), \tag{15}$$

that is $P_{\max}$ is a projection on the (z, P) plane. It is also found that at the maximum power condition, z is given by a power series in λ, namely,

$$z_P = \sqrt{\varepsilon} + \tfrac{1}{2}\left(1 - \sqrt{\varepsilon}\right)^2 \lambda + \tfrac{1}{4}\left(1 - \sqrt{\varepsilon}\right)^2 \left[\left(1 - \sqrt{\varepsilon}\right)^2 / 2\sqrt{\varepsilon} - \ln \varepsilon\right]\lambda^2 + O(\lambda^3) \tag{16}$$

Upon susbtituting Equation (16) in Equation (6) and because the terms in Equation (16) are positive, an upper bound for the efficiency is obtained when $\lambda = 0$, i.e., when the compression ratio $r_C = V_3/V_1$ goes to infinity, it results in the following:

$$\eta_{\max} = 1 - z_P\left(\lambda = 0\right) = \eta_{CAN}. \tag{17}$$

The equivalent of Equation (16) for the ecological function with this procedure was obtained in [12] by substituting Equation (9) in Equation (3), and the entropy production, $\sigma = \Delta S / t_{TOT}$. Using Equation (8) in the case $k = 1$, and the total time t_{TOT} given by Equations (10)–(13), the ecological function becomes

$$E = \alpha T_1 \frac{\left(1 + \varepsilon - 2z\right)\left(1 + \lambda \ln z\right)}{\frac{1}{1-u} + \frac{z}{zu - \varepsilon}}. \tag{18}$$

Upon maximizing the function $E = E(u, z)$ (ε is defined positive and λ is defined semi-positive, being external parameters), $\partial E/\partial u = 0$ and $\partial E/\partial z = 0$, for the first one $u = u(z)$ is as in case of

http://dx.doi.org/10.5772/61011

maximizing power output, and for the second one, the following relation between the variables z and u is obtained:

$$\left[2\left(1+\lambda\ln z\right)z-\lambda\left(1+\varepsilon-2z\right)\right]\left(z-\varepsilon\right)\left(zu-\varepsilon\right)=\left(1+\varepsilon-2z\right)\left(1+\lambda\ln z\right)\left(1-u\right)\varepsilon z. \tag{19}$$

The equation that z obeys at the maximum of the ecological function is obtained as follows:

$$\left[2\left(1+\lambda\ln z\right)z-\lambda\left(1+\varepsilon-2z\right)\right]\left(z-\varepsilon\right)=\left(1+\varepsilon-2z\right)\left(1+\lambda\ln z\right)\varepsilon. \tag{20}$$

We find, upon taking the implicit successive derivatives of Equation (20) with respect to λ, the following one-power series in λ :

$$\begin{aligned} z_E=\sqrt{\tfrac{1}{2}\left(\varepsilon+\varepsilon^2\right)}\Bigg\{1+\left[\tfrac{1}{4}\left(1+3\varepsilon\right)\sqrt{\tfrac{2}{\varepsilon+\varepsilon^2}}-1\right]\lambda+\left(\frac{\frac{1}{16}\left(1+3\varepsilon\right)}{\varepsilon+\varepsilon^2}-\frac{1}{2}\sqrt{\frac{2}{\varepsilon+\varepsilon^2}}\ln\sqrt{\tfrac{1}{2}\left(\varepsilon+\varepsilon^2\right)}\right)\times \\ \left(1+3\varepsilon-4\sqrt{\tfrac{1}{2}\left(\varepsilon+\varepsilon^2\right)}\right)\lambda^2+O\left(\lambda^3\right)\Bigg\} \end{aligned} \tag{21}$$

Furthermore, using Equation (21), we can write the efficiency as a power series in λ,

$$\eta_E\equiv 1-z_E\left(\varepsilon,\lambda\right). \tag{22}$$

When $\lambda=0$, the corresponding ecological efficiency with instantaneous adiabats is,

$$\eta_{EO}=1-z_{EO}\left(\varepsilon,\lambda=0\right)=1-\sqrt{\tfrac{1}{2}\left(\varepsilon+\varepsilon^2\right)}, \tag{23}$$

which is the maximum possible one for this operating regimen. From Equations (16) and (21) a linear approximation for the efficiency η in terms of compression ratio can be derived, $r_C=V_3/V_1$, and of the ratio T_C/T_H. It can be verified that $r_C\to\infty$ and $\lambda\to 0$ lead to the Curzon–Ahlborn–Novikov–Chambadal efficiency, now written as $\eta_{CAN}\equiv\eta_P(\lambda=0)=\eta_{PO}$. From Equation (16), the linear approximation can be obtained:

$$\eta_{PL}\left(\lambda\right)=1-\sqrt{\varepsilon}-\tfrac{1}{2}\left(1-\sqrt{\varepsilon}\right)^2\lambda, \tag{24}$$

and the corresponding linear approximation of ecological efficiency is as follows:

$$\eta_{EL}(\lambda) = 1 - \sqrt{\tfrac{1}{2}\left(\varepsilon^2 + \varepsilon\right)} - \left[\tfrac{1}{4}(1 + 3\varepsilon) - \sqrt{\tfrac{1}{2}\left(\varepsilon^2 + \varepsilon\right)}\right]\lambda. \tag{25}$$

As it is known in real compressors, the percent of volume in the total displacement of a piston into a cylinder is called the dead space ratio, and it is defined as $c = (\text{volume of dead space})/(\text{volume of displacement})$ [29]. In the Curzon–Ahlborn cycle, r_C appears as the reciprocal of c. It is found that $3\% \leq c \leq 10\%$; hence, $100/3 \geq r_C \geq 100/10$ or $33 > r_C \geq 10$. Supposing power plants working as a Curzon–Ahlborn cycle, a linear approximation of efficiency, Equations (24) and (25), values of efficiency appear around the experimental values. As an example, Table 1 shows a comparison between real values and linear approximation values, $\gamma = 1.67$, and the closeness of the linear approximation, in case of some modern power plants.

Nuclear power plant	$T_C\,(K)$	$T_H\,(K)$	η_{obs}	η_{EL} , $10 \leq r_C < 33$
Doel 4 (Belgium),	283	566	0.35000	0.37944–0.38224
Almaraz II (Spain)	290	600	0.34500	0.39234–0.39539
Sizewell B (UK)	288	581	0.36300	0.38277–0.38563
Cofrentes (Spain)	289	562	0.34000	0.36844–0.37103
Heysham (UK)	288	727	0.40000	0.46036–0.46506

Table 1. Comparison of experimental efficiencies with linear ecological approximation

2.2. Nonlinear heat transfer law

The ecological efficiency has been calculated using Dulong and Petit's heat transfer law in [30], maximizing ecological function for instantaneous adiabats. When the time for all the processes of the Curzon and Ahlborn cycle is taken into account, efficiencies in both regimens, maximum power output and maximum ecological function, can be obtained, following the procedure employed. Suppose an ideal gas as a working substance in a cylinder with a piston that exchanges heat with the reservoirs, and using a heat transfer law of the form

$$\frac{dQ}{dt} = \alpha\left(T_f - T\right)_i^k, \tag{26}$$

where $k > 1$, α is the thermal conductance assumed the same for both reservoirs, dQ/dt is the rate of heat Q exchange, and T_i and T_f are the temperatures for the heat exchange process. From the first law of thermodynamics and under mechanical equilibrium condition, i.e., $p = p_{ext}$, because the working substance is an ideal gas, $U = U(T)$, one obtains

http://dx.doi.org/10.5772/61011

$$\frac{dQ}{dt} = p\frac{dV}{dt} \text{ or } \alpha\left(T_f - T_i\right)^k = \frac{RT_i}{V}\frac{dV}{dt}. \tag{27}$$

Equation (27) implies that the times along the isothermal processes in Figure 1 are, respectively,

$$t_1 = \frac{RT_{HW}}{\alpha\left(T_H - T_{HW}\right)^k}\ln\frac{V_2}{V_1} \text{ and } t_3 = \frac{RT_{CW}}{\alpha\left(T_{CW} - T_C\right)^k}\ln\frac{V_3}{V_4} \tag{28}$$

The corresponding heat exchanged Q_H and Q_C become, respectively,

$$Q_H = RT_{HW}\ln\frac{V_2}{V_1} \text{ and } Q_C = RT_{CW}\ln\frac{V_4}{V_3}, \tag{29}$$

where R is the universal gas constant and V_1, V_2, V_3, V_4, are the corresponding volumes for the states 1, 2, 3, and 4 in Figure 1 also. The times of the adiabatic processes are assumed as

$$t_2 = \frac{RT_{HW}}{\alpha\left(T_H - T_{HW}\right)^k\left(\gamma - 1\right)}\ln\frac{T_{HW}}{T_{CW}}, \text{ and } t_4 = \frac{-RT_{CW}}{\alpha\left(T_{CW} - T_C\right)^k\left(\gamma - 1\right)}\ln\frac{T_{CW}}{T_{HW}} \tag{30}$$

where $\gamma \equiv C_P / C_V$ has been used. With these results, the form for the power output is

$$P = \frac{T_1^k\alpha\left(1-z\right)\left(1+\lambda\ln z\right)}{\frac{1}{\left(1-u\right)^k} + \frac{z}{\left(zu-\varepsilon\right)^k}}, \tag{31}$$

with the same used parameters. By means of $\partial P / \partial u = 0$ and $\partial P / \partial z = 0$, one obtains

$$u = \frac{z^{\frac{2}{k+1}} + \varepsilon}{z + z^{\frac{2}{k+1}}}, \tag{32}$$

and the resulting expression for the implicit function $z = z(\lambda, \varepsilon)$, for a given k,

$$\begin{aligned}\left[z^{\frac{2}{k+1}}\left(z-\varepsilon\right)\left(\lambda\left(1-z\right) - z\left(1+\lambda\ln z\right)\right) + zk\left(z^{\frac{2}{k+1}} + \varepsilon\right)\left(1-z\right)\left(1+\lambda\ln z\right)\right]\left(z^{\frac{2k}{k+1}} + z\right) \\ - z\left(1-z\right)\left(1+\lambda\ln z\right)\left[z^2 + \varepsilon z^{\frac{2k}{k+1}} + z^{\frac{2}{k+1}}\left(z-\varepsilon\right)\right] = 0.\end{aligned} \tag{33}$$

With reasonable approximations, only for the exponents in Equation (33), the following can be obtained:

$$(1+\lambda)\Big((k\varepsilon+zk)(1-z)-z(z-\varepsilon)\Big)+\lambda(1-z)(1-\varepsilon)-(1+\lambda\ln z)(1-z)z=0. \tag{34}$$

Equation (34) allows to the explicit expression for the function $z=z(\varepsilon,\ k)$ when $\lambda=0$,

$$z_{OP}(\varepsilon,k)=\frac{(k-1)(1-\varepsilon)\pm\sqrt{(\varepsilon-1)^2(1-k)^2+4k^2\varepsilon}}{2k}. \tag{35}$$

Taking now $k=5/4$ in Equation (35), one obtains the following value for the physically acceptable and approximated solution of Equation (33), namely,

$$z_{OPDP}=\frac{1-\varepsilon+\sqrt{\varepsilon^2+98\varepsilon+1}}{10}. \tag{36}$$

The numerical results for $\eta_{OPDP}=1-z_{OPDP}$ are compared with η_{CAN} and the observed efficiency, η_{obs}, which are in good agreement with the reported values. Now, assuming that z obtained from Equation (34) can be expressed as a power series in the parameter λ, the expression for the efficiency at maximum power output regimen is as follows:

$$\eta_{PDP}=1-z_{PDP}(\lambda,\varepsilon)=1-z_{OPDP}\Big[1+B_1(\varepsilon)\lambda+B_2(\varepsilon)\lambda^2+O(\lambda^3)\Big]. \tag{37}$$

One can find B_j, $j=1, 2, ...etc.$, through successive derivatives respect to λ. The first one is

$$B_1(\varepsilon)=\frac{16(1-z_{OPDP})(\varepsilon-z_{OPDP})}{z_{OPDP}(5-4\varepsilon-40z_{OPDP})} \tag{38}$$

Now, the ecological function for Curzon and Ahlborn engine takes the form

$$E(u,z)=\frac{T_H^k\alpha(1+\lambda\ln z)(1+\varepsilon-2z)}{\frac{1}{(1-u)^k}+\frac{z}{(zu-\varepsilon)^k}}. \tag{39}$$

We find the function $z(\varepsilon)$ from the maximization of function $E(u,\ z)$ and the efficiency for $k=5/4$. Upon setting $\partial E/\partial u=0$ and $\partial E/\partial z=0$, one obtains from the first condition that

http://dx.doi.org/10.5772/61011

$$u = \frac{z^{\frac{2}{k+1}} + \varepsilon}{z + z^{\frac{2}{k+1}}}, \tag{40}$$

and from the second one,

$$\frac{\left(\left(1+\varepsilon-2z\right)\lambda - 2z\left(1+\lambda\ln z\right)\right)\left(zu-\varepsilon\right)}{\left(1+\lambda\ln z\right)\left(1+\varepsilon-2z\right)\left(zu-\varepsilon-kuz\right)z} - \frac{\left(1-u\right)^k}{\left(zu-\varepsilon\right)^k + z\left(1-u\right)^k} = 0. \tag{41}$$

Substituting now Equation (40) for u in Equation (41), one obtains the following expression:

$$\begin{aligned}&\left(z^2 + z^{\frac{k+3}{k+1}}\right)\left(z-\varepsilon\right)\left(-2\left(1+\lambda\ln z\right)z + \left(1+\varepsilon-2z\right)\lambda\right) = \\ &= \left(z^{\frac{k+3}{k+1}} - \varepsilon z^{\frac{2}{k+1}} - \left(z^{\frac{k+3}{k+1}} + z\varepsilon\right)k\right)z\left(1+\lambda\ln z\right)\left(1+\varepsilon-2z\right).\end{aligned} \tag{42}$$

The analytical solution of Equation (42) is not feasible when the exponents of z are not integers, which is the present case, $k = 5/4$. The numerical solution of Equation (42) shows that anyone solution falls into the region bounded by solutions for $\lambda = 0$ and $\lambda = 1$ [28]. It can be appreciated that within the values of $0 \le \varepsilon \le 1$, which are the only physically relevant, the curve represented by Equation (42) can be fitted with a parabolic curve. The simplest approximation that allows for a parabolic fit for $0 \le \lambda \le 1$ modifying the exponents leads to the approximate analytical expression for $z(\varepsilon, \lambda)$ as

$$2\left(-2\left(1+\lambda\ln z\right)z + \left(1+\varepsilon-2z\right)\lambda\right)\left(z-\varepsilon\right) - \left(1+\lambda\ln z\right)\left(1+\varepsilon-2z\right)\left(\left(z-\varepsilon\right) - \left(z+\varepsilon\right)k\right) = 0. \tag{43}$$

For the case $\lambda = 0$, that is instantaneous adiabats, and with $k = 5/4$, Equation (43) becomes

$$z_{OEDP} = \frac{1 - \varepsilon + \sqrt{649\varepsilon^2 + 646\varepsilon + 1}}{36}. \tag{44}$$

Any other root has no physical meaning because efficiencies must always be positive. Adequate comparison between fitted numerical values of η_{MEDP} in [30] and $\eta_{OEDP} = 1 - z_{OEDP}$ is in [28] and later in [25]. Assuming z given by Equation (43) as a power series in the parameter λ, efficiency can be found as follows:

$$\eta_{EDP} = 1 - z_{EDP}\left(\lambda, \varepsilon\right) = 1 - z_{OEDP}\left[1 + b_1\left(\varepsilon\right)\lambda + b_2\left(\varepsilon\right)\lambda^2 + O\left(\lambda^3\right)\right]. \tag{45}$$

At last taking, $z_0=z_{EDP}(\varepsilon,\ \lambda=0)=z_{OEDP}(\varepsilon)$, and from Equation (43), coefficients are found by successively taking the derivative respect to λ and evaluating at $\lambda=0$. The first one leads to the linear approximation for ecological efficiency since Equation (45), as follows:

$$b_1(\varepsilon)=\frac{-2z_0+2\varepsilon-6z_0\varepsilon+2\varepsilon^2+4z_0^2}{z_0\left(-9z_0-\frac{1}{4}\varepsilon+\frac{1}{4}\right)}, \tag{46}$$

3. The non-endoreversible Curzon and Ahlborn cycle

By contrast, in finite time, thermodynamics is usually considered an endoreversible Curzon–Ahlborn cycle, but in nature, there is no endoreversible engine. Thus, some authors have analyzed the non-endoreversible Curzon and Ahlborn cycle. Particularly in [16] has been analyzed the effect of thermal resistances, heat leakage, and internal irreversibility by a non-endoreversibility parameter, advanced in [14],

$$I_S\equiv\frac{\Delta S_C}{\Delta S_H}, \tag{47}$$

where ΔS_C is the change of entropy during the exchange of heat from the engine to cold reservoir, and ΔS_H is the change of entropy during the exchange of heat from the hot reservoir to engine. The non-endoreversible Curzon–Ahlborn cycle is shown in Figure 2. The efficiency at maximum power output for instantaneous adiabats is

$$\eta_m=1-\sqrt{I_S\varepsilon},\quad I_S>1. \tag{48}$$

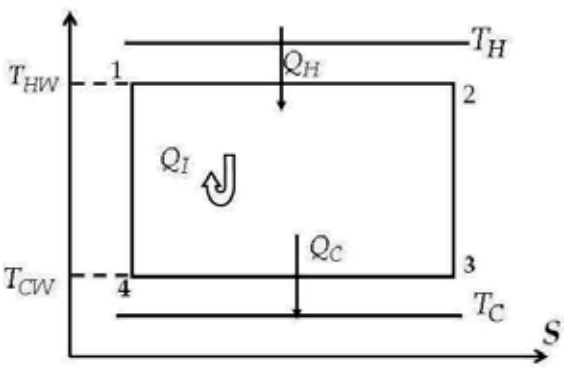

Figure 2. Curzon and Ahlborn cycle in the *S-T* plane. Q_I is a generated internally heat.

http://dx.doi.org/10.5772/61011

Following the procedure in [16], have been found expressions to measure possible reductions of undesired effects in heat engines operation [17], and has been pointed out that I_S is not dependent of ε and rewrote Equation (48) as

$$\eta_m = 1 - \sqrt{\frac{\varepsilon}{I}}, \quad I \equiv \frac{1}{I_S}, \quad 0 < I < 1. \tag{49}$$

Moreover, in [31] has been applied variational calculus showing that the saving function in [17] and modified ecological criteria are equivalent. In this section, internal irreversibilities are taken into account to obtain Equation (4), replacing $(\varepsilon^2 + \varepsilon)/2I$ instead $(\varepsilon^2 + \varepsilon)/2$ in case of a non-endoreversible Curzon and Ahlborn cycle. The procedure in [5] is combined with the cyclic model in [16] to obtain the form of power output function and of ecological function.

3.1. Curzon and Ahlborn cycle with instantaneous adiabats

Suppose a thermal engine working like a Curzon and Ahlborn cycle, in which an internal heat by internal processes of working fluid appears, assuming ideal gas as working fluid. The Clausius inequality with the parameter of non-endoreversibility becomes

$$I_S \frac{Q_H}{T_{HW}} - \frac{Q_C}{T_{CW}} = 0. \tag{50}$$

The changes of entropy are ΔS_C and ΔS_H during the heat exchange between the engine and its reservoirs. From Equation (50), $Q_C = (T_{CW}/T_{HW}) I_S Q_H$, and clearly $I_S \geq 1$. Thus, the heat exchanges between the thermal engine and its reservoirs are

$$Q_H = RT_{HW} \ln \frac{V_2}{V_1} \text{ and } Q_C = \frac{T_{CW}}{T_{HW}} I_S RT_{HW} \ln \frac{V_2}{V_1}. \tag{51}$$

The volumes in the states of change of process in the cycle are V_1, V_2, V_3, V_4, and the total made work by the engine can be written as

$$W_I = RT_{HW} \left(1 - I_S \frac{T_{CW}}{T_{HW}}\right) \ln \frac{V_2}{V_1}, \tag{52}$$

Assume the exchange of heat as Equation (8) with $k = 1$ and an internal generated heat Q_I. For reversible adiabatic processes, $TV^{\gamma-1} = constant$, with $\gamma = C_P/C_V$, so that $V_2/V_1 = V_3/V_4$ is obtained. For instantaneous adiabatic processes, the total time is

$$t_{TOT} = \frac{RT_{HW}}{\alpha}\left[\frac{1}{T_H - T_{HW}} + \frac{I_S}{T_{CW} - T_C}\cdot\frac{T_{CW}}{T_{HW}}\right]\ln\frac{V_2}{V_1}, \tag{53}$$

with the changes $Z_I = I_S T_{CW} / T_{HW}$ and $u = T_{HW} / T_H$, the power output is

$$P_I = \alpha T_1\left(1 - Z_I\right)\left(\frac{1}{1-u} + \frac{I_S Z_I}{Z_I u - \varepsilon I_S}\right)^{-1}. \tag{54}$$

Also, the variation of entropy in the cycle can be written as

$$\Delta S_I = -\frac{Q_H}{T_H} + \frac{Q_C}{T_C} = -R\frac{T_H}{T_C}\left(\varepsilon - Z_I\right)\ln\frac{V_3}{V_1}, \tag{55}$$

and Equation (18) is modified as

$$E_I = \alpha T_1\left(1 - 2Z_I + \varepsilon\right)\left(\frac{1}{1-u} + \frac{I_S Z_I}{Z_I u - \varepsilon I_S}\right)^{-1}. \tag{56}$$

Now, the following is obtained from the conditions $\partial P / \partial u = 0$ and $\partial P / \partial Z_I = 0$:

$$u = \left(Z_I + \varepsilon\sqrt{I_S}\right)\sqrt{I_S}\left[Z_I\left(1 + \sqrt{I_S}\right)\right]^{-1}, \tag{57}$$

and a physically possible solution for Z_I is found, which leads to the efficiency

$$\eta_{CANI} = 1 - \sqrt{\frac{\varepsilon}{I}}, \tag{58}$$

when the change $I_S = 1/I$ proposed in [17] is used. Similar results can be obtained for the ecological function. Thus, from Equation (56) for the same variables u and Z_I, function $u = u(z)$ is obtained as Equation (57), and the physically possible solution of Z_I leads to ecological efficiency as in [23],

$$\eta_{EI} = 1 - \sqrt{\frac{\varepsilon^2 + \varepsilon}{2I}}, \tag{59}$$

For the suitable values of parameter $I = 1/I_S$, found in [17] as $0.8 \le I \le 0.9$ 0, Table 2 shows the values of the ecological efficiency, Equation (59), compared with the experimental values of

the efficiency, η_{obs}. The intervals of values of the efficiency are improved in the sense that they are nearer to the reported experimental values in literature.

Power Plant	$T_2(K)$	$T_1(K)$	η_{obs}	η_{EI}
Doel 4 (Belgium), 1985	283	566	0.35000	0.31535 to 0.3545
Almaraz II, Spain	290	600	0.34500	0.3306 to 0.36889
Sizewell B, UK	288	581	0.36300	0.3198 to 0.35821
Cofrentes, Spain	289	562	0.34000	0.30238 to 0.34228
Heysham, UK	288	727	0.40000	0.41206 to 0.44568

Table 2. Comparison of experimental efficiencies with efficiencies from Equation (25)

3.2. Curzon–Ahlborn cycle with noninstantaneous adiabats

In order to include the compression ratio in the analysis of Curzon and Ahlborn cycle, it is necessary to suppose finite time for the adiabatic processes. Hence, as it is known, with ideal gas as working fluid and using the Newton heat transfer law, the following can be written:

$$\frac{dQ}{dt} = p\frac{dV}{dt}, \tag{60}$$

and because $p = RT/V$, Equation (60) is now

$$\frac{dQ}{dt} = \frac{RT}{V}\frac{dV}{dt} = RT\frac{d}{dt}\left(\ln V\right). \tag{61}$$

Then again, internal energy U depends only on the initial and final states, so the adiabatic expansion in the cycle can be written as

$$\frac{1}{V}\frac{dV}{dt} = \alpha\frac{T_H - T_{HW}}{RT_{HW}}. \tag{62}$$

The integration of Equation (62) leads to the time of the adiabatic expansion in the cycle,

$$t_2 = \left(RT_{HW}\ln\tfrac{V_3}{V_2}\right)\left[\alpha(T_H - T_{HW})\right]^{-1}, \tag{63}$$

and taking into account the form that acquires the yielded heat Q_C based on the absorbed heat Q_H, Equation (51), the time of the adiabatic processes can be assumed as

$$t_{ad} = \left(RT_{HW} \ln \tfrac{V_3}{V_4}\right)\left[\alpha\left(T_H - T_{HW}\right)\right]^{-1} + \left[RT_{HW}(I_S \tfrac{T_{CW}}{T_{HW}}) \ln \tfrac{V_1}{V_4}\right] \cdot \left[\alpha\left(T_{CW} - T_C\right)\right]^{-1}, \tag{64}$$

and the total time of the non-endoreversible cycle is as follows:

$$t_{TOT}(ad) = \frac{RT_{HW}}{\alpha}\left[\tfrac{1}{T_H - T_{HW}} + \tfrac{I_s}{T_{CW} - T_C} \cdot \tfrac{T_{CW}}{T_{HW}}\right] \ln \tfrac{V_3}{V_1}, \tag{65}$$

So that a new expression for power output in the cycle using the changes of variables in Equation (54) and Equation (56) is found, namely,

$$P_{I\lambda} = \alpha T_H \left(1 - Z_I\right)\left(1 + \lambda \ln Z_I - \lambda \ln I_S\right) \cdot \left[\tfrac{1}{1-u} + \tfrac{Z_I I_S}{Z_I u - \varepsilon I_S}\right]^{-1}, \tag{66}$$

with $\lambda \equiv 1/((\gamma - 1)\ln r_C)$, and the compression ratio is $r_C \equiv V_3/V_1$. The entropy production with the same changes of variables is found, and the new expression for ecological function is

$$E_{I\lambda} = \alpha T_H \left(1 - 2Z_I + \varepsilon\right)\left(1 + \lambda \ln Z_I - \ln I_S\right) \cdot \left[\tfrac{1}{1-u} + \tfrac{I_S Z_I}{Z_I u - \varepsilon I_S}\right]^{-1}. \tag{67}$$

In order to maximize power output, Equation (66), the conditions $\partial P_{I\lambda}/\partial u = 0$ and $\partial P_{I\lambda}/\partial Z_I = 0$ are necessary. Also, in order to maximize ecological function, Equation (67), the conditions $\partial E_{I\lambda}/\partial u = 0$ and $\partial E_{I\lambda}/\partial Z_I = 0$ are necessary. Hence, one can find the form of Z_I for each maximized features function. In case of maximizing power output, the following is obtained:

$$Z_I^2 + \lambda Z_I^2 \ln Z_I - \lambda Z_I^2 \ln I_S + \lambda \varepsilon I_S - \lambda Z_I - \lambda Z_I \varepsilon I_S + \lambda Z_I^2 - \varepsilon I_S - \lambda \varepsilon I_S \ln Z_I + \lambda \varepsilon I_S \ln I_S = 0, \tag{68}$$

and in case of maximizing ecological function, the following is obtained:

$$2Z_I^2\left(1 - \lambda \ln I_S + \lambda\right) - \lambda Z_I\left(\varepsilon + 1 + 2\varepsilon I_S\right) + 2\lambda Z_I^2 \ln Z_I = \left(\varepsilon + 1\right)\left(1 - \lambda - \lambda \ln I_S + \lambda \ln Z_I\right)\varepsilon I_S. \tag{69}$$

When $\lambda = 0$ and $I_S = 1$ the corresponding expressions shown in [5] and in [12] for maximum power output and maximum ecological function are recovered. Moreover, expressions of the Curzon–Ahlborn–Novikov–Chambadal efficiency and the ecological efficiency found in [1] and [4] are recovered. Non instantaneous adiabats imply $\lambda \neq 0$, and Z_I can be expanded in a power series of λ. The simplest expansion is the linear approach, so if Z_I is written for each case of objective function, one can obtain, respectively,

$$Z_{PI} = a_0 + a_1\lambda + \cdots \text{ and } Z_{EI} = b_0 + b_1\lambda + \cdots. \tag{70}$$

http://dx.doi.org/10.5772/61011

Parameters a_i and b_i can be calculated as in [5] and in [12]. They are small and greater than zero. They go to zero as $i \to \infty$; thus, it is possible to ensure the convergence of series. Linear approximation only requires finding a_0 and a_1 (or b_0 and b_1), which are, in case of maximum power output, as follows:

$$a_0 = \sqrt{\varepsilon I_S} \text{ and } a_1 = \tfrac{1}{2}\left(1 - \sqrt{\varepsilon I_S}\right), \tag{71}$$

and in case of maximum ecological function,

$$b_0 = \sqrt{\tfrac{1}{2} I_S \left(\varepsilon^2 + \varepsilon\right)} \text{ and } b_1 = \tfrac{1}{4}\left(1 + \varepsilon + 2\varepsilon I_S\right) - \sqrt{\tfrac{1}{2} I_S \left(\varepsilon^2 + \varepsilon\right)}. \tag{72}$$

The linear approximation of efficiency, at maximum power output and at maximum ecological function, can now be derived from Equation (71) or Equation (72), respectively, as

$$\eta_{CANL} = 1 - Z_{PI} \text{ or } \eta_{EL} = 1 - Z_{EI}. \tag{73}$$

It is important to note that compression ratio has no arbitrary values, as discussed in Section 2.1. Thus, for $r_C = 10$ and the extreme values of the range of values for I, $I = 1/I_S$, found in [17] for real engines with a gas as working fluid, namely, $I = 0.8$ and $I = 0.9$, the efficiency is obtained as a function of the parameter ε.

4. Stirling and Ericsson cycles

As it is known, the thermal engines can be endothermic or exothermic. Among the first engines, the best known are Otto and Diesel, and among the second two engines, very interesting and similar to the theoretical Carnot engine are Stirling and Ericsson engines [32,33]. In particular, a Stirling engine is a closed-cycle regenerative engine initially used for various applications, and until the middle of last century, they were manufactured on a large scale. However, the development of internal combustion engines from the mid-nineteenth century and the improvement in the refining of fossil fuels influenced the abandonment of the Stirling and Ericsson engines in the race for industrialization, gradually since the early twentieth century. Reference [34] is an interesting paper devoted to Stirling engine.

In the classical equilibrium thermodynamics, Stirling and Ericsson cycles have an efficiency that goes to the Carnot efficiency, as it is shown in some textbooks. These three cycles have the common characteristics, including two isothermal processes. The objection to the classical point of view is that reservoirs coupled to the engine modeled by any of these cycles do not have the same temperature as the working fluid because this working fluid is not in direct thermal contact with the reservoir. Thus, an alternative study of these cycles is using finite

time thermodynamics. Thus, since the end of the previous century, and on recent times, the characteristics of Stirling and Ericsson engines have resulted in renewed interest in the study and design of such engines, and in the analysis of its theoretical idealized cycle, as it is shown in many papers, [22,35-37] among others. Nevertheless, the discussion on these engines and its theoretical model has not been exhausted.

In this section, an analysis of the Stirling and Ericsson cycles from the viewpoint of finite time thermodynamics is made. The existence of finite time for heat transfer in isothermal processes is proposed, but the cycles are analyzed assuming they are not endoreversible cycles, through the factor that represents their internal irreversibilities [14], so that the proposed heat engine model is closer to a real engine. Some results in reference [22] are used, and a methodology to obtain a linear approximation of efficiency, including adequate parameters, is shown. Variable changes are made right, like those used in [5] and in [23,25]. This section is a summary of obtained results in [38].

4.1. Stirling cycle

Now, as it is known, Stirling cycle consists of two isochoric processes and two isothermal processes. At finite time, the difference between the temperatures of reservoirs and the corresponding operating temperatures is considered, as shown in Figure 3. To construct expressions for power output and ecological function for this cycle, some initial assumptions are necessary. First, the heat transfer is supposed as Newton's cooling law for two bodies in thermal contact with temperatures T_i and T_f, $T_i > T_f$, with a rapidity of heat change dQ/dt, and a constant thermal conductance α, which for convenience is assumed to be equal in all cases of heat transfer as follows:

$$\frac{dQ}{dt} = \alpha\left(T_i - T_f\right). \tag{74}$$

On the other hand, it is assumed that the internal processes of the system cause irreversibilities that can be represented by the factor I_S previously presented, so from the second law of thermodynamics, the following can be written:

$$Q_C = \frac{T_{CW}}{T_{HW}} I_S Q_H. \tag{75}$$

Power output is defined as

$$P = \frac{Q_H - Q_C}{t_{TOT}}. \tag{76}$$

http://dx.doi.org/10.5772/61011

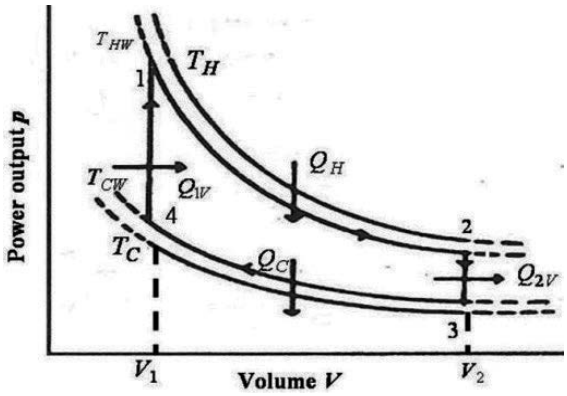

Figure 3. Idealized Stirling cycle at the *V–p* (volume vs pressure) plane.

With ideal gas as working substance for an isothermal process, the equation of state leads to

$$\frac{RT_i}{V}\frac{dV}{dt} = \alpha\left(T_i - T_f\right). \tag{77}$$

An assumption for the cycle is that heating and cooling at constant volume is performed as

$$\left|\frac{dT}{dt}\right| = r_V = constant, \tag{78}$$

where it is not difficult to show that it meets

$$\left|Q_{1V}\right| = \left|Q_{2V}\right|. \tag{79}$$

By contrast, from the equilibrium conditions, it can be assumed

$$\frac{dU}{dt} = \frac{dQ}{dt} = C_V\frac{dT}{dt} = r_V C_V, \tag{80}$$

and the heating and cooling, respectively, from the first law of thermodynamics are

$$Q_{1V} = C_V\left(T_{HW} - T_{CW}\right) = \Delta U_{41} \text{ and } Q_{2V} = C_V\left(T_{CW} - T_{HW}\right) = \Delta U_{23}, \tag{81}$$

and the time for each isochoric processes is given as

$$t_V = \frac{1}{r_V}\left(T_{HW} - T_{CW}\right). \tag{82}$$

The time for the isothermal processes can be found from Equation (77) as

$$t_1 = \frac{RT_{HW}}{\alpha\left(T_H - T_{HW}\right)}\ln\frac{V_2}{V_1} \text{ and } t_2 = -\frac{RT_{CW}}{\alpha\left(T_{CW} - T_C\right)}\ln\frac{V_1}{V_2}. \tag{83}$$

The negative sign in t_2 is just because there is no negative time, the total time of cycle is

$$t_{TOT} = \frac{RT_{HW}}{\alpha\left(T_H - T_{HW}\right)}\ln\frac{V_2}{V_1} - \frac{RT_{CW}}{\alpha\left(T_{CW} - T_C\right)}\ln\frac{V_1}{V_2} + \frac{2}{r_V}\left(T_{HW} - T_{CW}\right). \tag{84}$$

Since its definition and taking into account Equation (76), the power output of cycle is written as

$$P_{SI} = \frac{Q_H - I_S\frac{T_{CW}}{T_{HW}}Q_H}{t_1 + t_2 + 2t_V}. \tag{85}$$

Now, with the change of variables used in the previous section in Equations (54) and (56), and taking into account the ratio of temperatures of the heat reservoirs, used in Equations (1) and (2), with the parameter $\lambda = [(\gamma - 1)\ln(V_2/V_1)]^{-1}$ that includes the compression ratio of cycle, in this case $V_2/V_1 = r_C$, the power output of Stirling cycle takes the form

$$P_{SI} = \frac{\alpha T_H\left(1 - Z_I\right)}{\frac{1}{1-u} + \frac{I_S Z_I}{Z_I u - I_S\varepsilon} + \frac{2\alpha T_H}{C_V I_S r_V}\left(I_S - Z_I\right)\lambda}. \tag{86}$$

The optimization conditions $\left(\partial P_{SI}/\partial u\right)_{Z_I=const} = 0$ and $\left(\partial P_{SI}/\partial Z_I\right)_{u=const} = 0$ permit find the function $Z_I = Z_I(\varepsilon, I_S, \lambda)$. From the first one, $u = u(Z_I, I_S)$ is obtained as

$$u = \frac{\left(Z_I + \varepsilon\sqrt{I_S}\right)\sqrt{I_S}}{Z_I\left(1 + \sqrt{I_S}\right)}, \tag{87}$$

and from the second one, a solution physically adequate Z_{IP} can be obtained by

http://dx.doi.org/10.5772/61011

$$\frac{\left(1+\sqrt{I_S}\right)^2 r_V C_V I_S + 2\alpha T_H \lambda\left(I_S - 2Z_I + \varepsilon I_S\right)}{-\left(Z_I - \varepsilon I_S\right) + \left(1 - Z_I\right)} = \\ = \frac{Z_I\left(1+\sqrt{I_S}\right)^2 r_V C_V I_S + 2\alpha T_H \lambda\left(Z_I - \varepsilon I_S\right)\left(I_S - Z_I\right)}{\left(1 - Z_I\right)\left(Z_I - \varepsilon I_S\right)}. \tag{88}$$

Thus, that the efficiency at maximum power output can be written as

$$\eta_{SIP} = 1 - Z_{IP}\left(\varepsilon, I_S, \lambda\right). \tag{89}$$

For known values of parameters C_V, α, and T_H, in the limit $\lambda \to 0$, namely, $V_2/V_1 \to \infty$, the efficiency of non-endoreversible Stirling cycle, η_{SIP}, goes to the efficiency for the non-endoreversible Curzon and Ahlborn cycle, as can be seen from Equation (86),

$$\eta_{SIP} \to \eta_m = 1 - \sqrt{\tfrac{\varepsilon}{I}}. \tag{90}$$

The analysis for ecological function is similar to power output, and also leads to similar results. The shape of function $u = u(Z_I, I_S, \varepsilon)$ is the same as in Equation (87), but the form of $Z_I = Z_I(\varepsilon, I_S, \lambda)$ changes. Because heating and cooling in both isochoric and isobaric processes are considered constant, and taking into account Equations (75) and (78), the change of entropy can be taken only for isothermal processes. Then, the change of entropy for the non-endoreversible cycle considered is

$$\Delta S = -\frac{Q_H}{T_H} + \frac{Q_C}{T_C} = -\frac{Q_H}{T_H} + I_S \frac{T_{CW}}{T_{HW}} \frac{Q_H}{T_C}, \tag{91}$$

which leads to the ecological function as

$$E_{SI} = \frac{Q_H}{t_{TOT}}\left(1 - 2I_S \frac{T_{CW}}{T_{HW}} + \frac{T_C}{T_H}\right) = \frac{RT_{HW}}{t_{tot}}\left(1 - 2I_S \frac{T_{CW}}{T_{HW}} + \frac{T_C}{T_H}\right)\ln\tfrac{V_2}{V_1}, \tag{92}$$

Where t_{TOT} is as Equation (84). With the same parameters definite in the previous section, ecological function can be written now as

$$E_{SI} = \frac{\alpha T_H\left(1 - 2Z_I + \varepsilon\right)}{\frac{1}{1-u} + \frac{I_S Z_I}{Z_I u - I_S \varepsilon} + \frac{2\alpha T_H}{C_V I_S r_V}\left(I_S - Z_I\right)\lambda}. \tag{93}$$

As in the case of power output, in order to find the efficiency at maximum ecological function, there are two conditions, namely, $(\partial E_{SI}/\partial u)_{Z_I=const}=0$ and $(\partial E_{SI}/\partial Z_I)_{u=const}=0$. These conditions lead to obtaining the parameter u as in Equation (87) and also $Z_{IE}=Z_{IE}(\varepsilon,\ I_S,\ \lambda)$ as an adequate solution for the second condition by the relation,

$$\begin{aligned}&\frac{Z_I\left(1+\sqrt{I_S}\right)^2 r_V C_V I_S+2T_H\alpha\lambda\left(Z_I-\varepsilon I_S\right)\left(I_S-Z_I\right)}{\left(1-2Z_I+\varepsilon\right)\left(Z_I-\varepsilon I_S\right)}=\\&=\frac{\left(1+\sqrt{I_S}\right)^2 r_V C_V I_S+2T_H\alpha\lambda\left(I_S-2Z_I+\varepsilon I_S\right)}{\left(1-2Z_I+\varepsilon\right)-2\left(Z_I-\varepsilon I_S\right)}\end{aligned}\tag{94}$$

The efficiency for the Stirling cycle at maximum ecological function can be written now as

$$\eta_{SIE}=1-Z_{IE}\left(\varepsilon,I_S,\lambda\right),\tag{95}$$

and $\lambda\to 0$ implies η_{SIE} goes to the efficiency for the non-endoreversible Curzon–Ahlborn cycle,

$$\eta_{SIE}\to\eta_{EI}=1-\sqrt{\tfrac{\varepsilon+\varepsilon^2}{2I}}.\tag{96}$$

The existence of a finite heat transfer in the isothermal processes is affected with the assumption of a non-endoreversible cycle with ideal gas as working substance. Power output and ecological function have also an issue that shows direct dependence on the temperature of the working substance. Expressions obtained with the changes of variables have the virtue of leading directly to the shape of the efficiency through Z_I function. Thus, in classical equilibrium thermodynamics, the Stirling cycle has its efficiency like the Carnot cycle efficiency; in finite time thermodynamics, this cycle has an efficiency in their limit cases as the Curzon–Ahlborn cycle efficiency.

4.2. Ericsson cycle

The Ericsson cycle consisting of two isobaric processes and two isothermal processes is shown in Figure 4. Now, it follows a similar procedure as in the Stirling cycle case. Thus, the hypothesis on constant heating and cooling, now at constant pressure, is expressed as

$$\left|\frac{dT}{dt}\right|=r_p=\text{constant}.\tag{97}$$

It is true that

http://dx.doi.org/10.5772/61011

$$\left|Q_{1p}\right|=\left|Q_{2p}\right|. \tag{98}$$

The equilibrium condition now is

$$\frac{dU}{dt}=C_p\frac{dT}{dt}-p\frac{dV}{dt}=r_pC_p-p\frac{dV}{dt}, \tag{99}$$

and the time for a constant pressure process is given as

$$t_p=\frac{T_{HW}}{r_p}\left(1-\frac{T_{CW}}{T_{HW}}\right). \tag{100}$$

The time for the isothermal processes can also be obtained from Equation (77) and can be written as

$$t_1=\frac{RT_{HW}}{\alpha\left(T_H-T_{HW}\right)}\ln\frac{V_2}{V_1}, t_2=-\frac{RT_{CW}}{\alpha\left(T_{CW}-T_C\right)}\ln\frac{V_4}{V_3}, \tag{101}$$

and the total time of cycle is now

$$t_{TOT}=\frac{RT_{HW}}{\alpha(T_H-T_{HW})}\ln\frac{V_2}{V_1}-\frac{RT_{CW}}{\alpha(T_{CW}-T_C)}\ln\frac{V_4}{V_3}+\frac{2}{r_p}(T_{HW}-T_{CW}), \tag{102}$$

so the power output of cycle from its definition and taking into account Equation (76) remains

$$P=\frac{Q_H-I_S\frac{T_{CW}}{T_{HW}}Q_H}{t_1+t_2+2t_p}. \tag{103}$$

With the change of variables used in the previous section, now the expression for the power output of the non-endoreversible Ericsson cycle is

$$P_{EI}=\frac{\alpha T_H\left(1-Z_I\right)}{\frac{1}{1-u}+\frac{Z_I}{Z_Iu-I_S\varepsilon}+\frac{2\alpha T_H}{C_VI_Sr_p}\left(I_S-Z_I\right)\lambda}, \tag{104}$$

which is essentially found for the Stirling cycle, with factor r_p instead of r_v. For extreme conditions, $\left(\partial P_{EI}/\partial u\right)_{u=const}=0$ and $\left(\partial P_{EI}/\partial u\right)_{Z_I=const}=0$ are obtained again using Equation (87), allowing us to find a physically acceptable solution $Z_{IP}=Z_{IP}(\varepsilon, I_S, \lambda)$ by

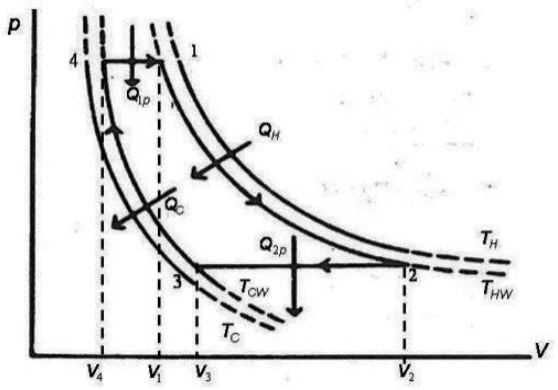

Figure 4. Idealized Ericsson cycle at the $V-p$ (volume vs pressure) plane.

$$\frac{\left(1+\sqrt{I_S}\right)^2 r_p C_V I_S + 2\alpha T_H \lambda\left(I_S - 2Z_I + \varepsilon I_S\right)}{-\left(Z_I - \varepsilon I_S\right)+\left(1-Z_I\right)} = \\ = \frac{Z_I\left(1+\sqrt{I_S}\right)^2 r_p C_V I_S + 2\alpha T_H \lambda\left(Z_I - \varepsilon I_S\right)\left(I_S - Z_I\right)}{\left(1-Z_I\right)\left(Z_I - \varepsilon I_S\right)} \tag{105}$$

Thus, at maximum power output regimen, the efficiency of non-endoreversible Ericsson cycle is

$$\eta_{EIP} = 1 - Z_{IP}\left(\varepsilon, I_S, \lambda\right). \tag{106}$$

The analysis for the case of ecological function is similar to the case of power output and also leads to similar results. The shape of the function $u=u(Z_I,\ I_S,\ \varepsilon)$ is the same as in Equation (87), but the form of $Z_I=Z_I(\varepsilon,\ I_S,\ \lambda)$ changes. Thus, because heating and cooling in isobaric processes are considered constant, the change of entropy can be taken only for the isothermal processes. Hence, for the non-endoreversible Ericsson cycle considered, we have

$$\Delta S = -\frac{Q_H}{T_H} + \frac{Q_C}{T_C} = -\frac{Q_H}{T_H} + I_S \frac{T_{CW}}{T_{HW}} \frac{Q_H}{T_C}, \tag{107}$$

from which the ecological function for the Ericsson cycle can be written as

$$E_{EI} = \frac{\alpha T_H\left(1-2Z_I+\varepsilon\right)}{\frac{1}{1-u} + \frac{Z_I}{Z_I u - I_S \varepsilon} + \frac{2\alpha T_H}{C_V I_S r_p}\left(I_S - Z_I\right)\lambda}, \tag{108}$$

where the parameter r_p takes the adequate value depending on the cycle analyzed. As in the case of power output, there are two conditions for maximum ecological function, namely,

http://dx.doi.org/10.5772/61011

$(\partial E_{EI}/\partial u)_{z_I=const}=0$ and $(\partial E_{EI}/\partial Z_I)_{u=const}=0$. These conditions lead to obtain parameter u as in Equation (87) and also $Z_{EI}=Z_{EI}(\varepsilon,\ I_S,\ \lambda)$ by

$$\frac{Z_I\left(1+\sqrt{I_S}\right)^2 r_p C_V I_S+2T_H\alpha\lambda\left(Z_I-\varepsilon I_S\right)\left(I_S-Z_I\right)}{\left(1-2Z_I+\varepsilon\right)\left(Z_I-\varepsilon I_S\right)}=$$
$$=\frac{\left(1+\sqrt{I_S}\right)^2 r_p C_V I_S+2T_H\alpha\lambda\left(I_S-2Z_I+\varepsilon I_S\right)}{\left(1-2Z_I+\varepsilon\right)-2\left(Z_I-\varepsilon I_S\right)} \tag{109}$$

The efficiency for Ericsson cycle at maximum ecological function can be written now as

$$\eta_{EIE}=1-Z_{IE}\left(\varepsilon,I_S,\lambda\right). \tag{110}$$

5. Concluding remarks

The developed methodology leads directly to appropriate expressions of the objective functions simplifying the optimization process. This methodology shows the consequences of assuming non-endoreversible cyle in the process of isothermal heat transfer through the factor $I_S=1/I$, which represents the internal irreversibilities of cycle, so that the proposed heat engine model is closer to a real engine. By contrast, as the known Carnot theorem provided a level of operation of heat engines, the Curzon and Ahlborn cycle provides levels of operation of such engines closer to reality. In this sense, the same manner within the context of classical equilibrium thermodynamics shows that in any cycle formed by two isothermal processes and any other pair of the same processes (isobaric, isochoric, and adiabatic), efficiency tends to Carnot cycle efficiency. In the context of finite time thermodynamics, any cycle as previously mentioned has an efficiency, which tends to Curzon and Ahlborn cycle efficiency. The above statements are independent if the cycle is considered endoreversible or non-endoreversible.

Acknowledgements

The authors thank the total support of the Universidad Autónoma Metropolitana (México).

Author details

Delfino Ladino-Luna*, Ricardo T. Páez-Hernández and Pedro Portillo-Díaz

*Address all correspondence to: dll@correo.azc.uam.mx

Universidad Autónoma Metropolitana-A, México

References

[1] Curzon, F.L., & Ahlborn, B. (1975). Efficiency of a Carnot engine at maximum power output. *Am. J. Phys.,* Vol. 43, pp 22–42.

[2] Chambadal, P. (1957). *Récupération de chaleur 'a la sortie d'un reactor*, chapter 3, Armand Colin (Ed.), Paris, France.

[3] Novikov, I.I. (1958). The efficiency of atomic power stations (a review). *J. Nucl. Energy II.*, Vol. 7, pp 125–128.

[4] Angulo-Brown, F. (1991). An ecological optimization criterion for finite-time heat engines. *J. Appl. Phys.,* Vol. 69, pp 7465–7469.

[5] Gutkowics-Krusin, D., Procaccia, I., & Ross, J. (1978) On the efficiency of rate processes. Power and efficiency of heat engines. *J. Chem. Phys.,* Vol. 69, pp 3898–3906.

[6] De Vos, A. (1985). Efficiency of some heat engines at maximum-power condition. *Am. J. Phys.* Vol. 53, pp 570–573.

[7] Agrawal, D.C., Gordon, J.M., & Huleihil, M. (1994). Endoreversible engines with finite-time adiabats. *Indian J. Eng. Mat. Sci*., Vol. 1, pp 195–198.

[8] Torres, J.L. (1988). Minimal rate of entropy as a criterion of merit for thermal engines. *Rev. Mex. Fis*., Vol. 34, pp 18–24.

[9] Angulo-Brown, F. (1991) An entropy production approach to the Curzon and Ahlborn cycle. *Rev. Mex. Fís*., Vol. 37, pp 87–96.

[10] Bejan, A. (1996). Entropy generation minimization: the new thermodynamics of finite-size devices and finite-time processes. *J. Appl. Phys*.., Vol. 79, pp 1191–1218.

[11] Cheng, Ch.Y. (1997) The ecological optimization of an irreversible Carnot heat engine. *J. Phys. D. Appl. Phys.,* Vol. 30, pp 1602–1609.

[12] Ladino-Luna, D., & de la Selva, S.M.T. (2000). The ecological efficiency of a thermal finite time engine. *Rev. Mex. Fís*., Vol. 46, pp 52–56.

[13] Chen, L., Zhou, J., Sun, F., & Wu, C. (2004). Ecological optimization of generalized irreversible Carnot engines. *Applied Energy*, Vol. 77, pp 327–338

[14] Ibrahim, O.M., Klein, S.A., & Mitchel, J.W. (1991). Optimum heat power cycles for specified boundary conditions. *Trans. ASME*., Vol. 113, pp 514–521.

[15] Wu, C., & Kiang, R.L. (1992). Finite-time thermodynamic analisys of a Carnot engine with internal irreversibility. *Energy*., Vol. 17, pp 1173–1178.

[16] Chen, L., Zhou, J., Sun, F., & Wu, C. (2004). Ecological optimization of generalized irreversible Carnot engines. *Applied Energy*, Vol. 77, pp 327–338.

[17] Velasco, S., Roco, J.M.M., Medina, A., & White, J.A. (2000). Optimization of heat engines including the saving of natural resources and the reduction of thermal pollution. *J. Phys. D,* Vol. 33, pp 355–359.

[18] Ust, Y., Sahim, B., & Sogut, O.S. (2005). Performance analysis and optimization of an irreversible dual-cycle based on an ecological coefficient of performance criterion. *Appl. En*., Vol. 8, pp 23–39.

[19] Wang, H., Liu S., & He J. (2009). Performance analysis and parametric optimum criteria of a quantum Otto heat engine with heat transfer effects. *Appl. Thermal Eng.,* Vol. 29, pp 706–711.

[20] Gordon, J.M., & Huleihil, M. (1992) General performance of real heat engines. *J. Appl. Phys.,* Vol. 7, pp 829–837.

[21] Ladino-Luna, D. (2010). Analysis of behavior of a Carnot type cycle. *Inf. Tecnológica*, Vol. 21, pp 79–86.

[22] Angulo-Brown, F., & Ramos-Madrigal, G. (1990). Sobre ciclos politrópicos a tiempo finito (On politropic cycles at finite time). *Rev. Mex. Fís.*; Vol. 39, pp 363–375.

[23] Ladino-Luna, D. (2007). On optimization of a non-endoreversible Curzon–Ahlborn cycle. *Entropy,* Vol. 9, pp 186–197.

[24] Ladino-Luna, D. (2011). Linear approximation of efficiency for a non-endoreversible cycle. *J. Energy Instit.,* Vol. 84, pp 61–65.

[25] Ladino-Luna, D., & Paéz-Hernández, R.T. (2011). Non-instantaneous adiabats in finite time. In: Thermodynamics—Physical Chemistry of Aqueous Systems. InTech (Ed), Rijeka, Croatia.

[26] Salamon, P., Andresen, B., & Berry, R.S. (1976). Thermodynamics in finite time I. Potencials for finite-time processes. *Phys. Rev. A,* Vol. 15, pp 2094–2101.

[27] Rubin, M. (1980). Optimal configuration of an irreversible heat engine with fixed compression ratio. *Phys. Rev. A.* Vol. 22, pp 1741–1752.

[28] Ladino-Luna, D. (2008). Approximate ecological efficiency with a non-linear heat transfer. *J. Energy Inst.,* Vol. 81, pp 114–117.

[29] Burghardt, M.D. (1982). *Engineering thermodynamics,* section 5.2. Harper and Row (Ed.), New York, Unites States of America.

[30] Angulo-Brown, F., & Páez-Hernández, R. (1993). Endoreversible thermal cycle with a non linear heat transfer law. *J. Appl. Phys.,* Vol. 74, pp 2216–2219.

[31] Angulo-Brown, F., Ares de Parga, G., & Árias-Hernández, L.A. (2002). A variational approach to ecological-type optimization criteria for finite-time thermal engine models. *J. Phys. D. Appl.Phys.*, Vol. 35, pp 1089–1093.

[32] Zemansky, M.W. (1968). *Heat and Thermodynamics*, chap. 7, McGraw-Hill Book Company (Ed.), New York, United States of America.

[33] Wark, K. (1983). *Thermodynamics,* chap. 16, McGraw-Hill Book Company (Ed.), United States of America.

[34] De la Herrán, J. (2005). El motor Stirling, reto a la tecnología (Stirling engine, technology challenge). *Información Científica y Tecnológica*, Vol. 3, pp 4–11.

[35] Kaushik, S. Ch., Tyagi, S.K., Bose, S.K., & Singhal, M.K. (2002). Performance evaluation of irreversible Stirling and Ericsson heat pump cycles. *Int. J. Therm. Sci*. Vol. 41, pp 193–200.

[36] Tyagi, S.K., Kaushik, S.K., & Salohtra, R. (2002). Ecological optimization and performance study of irreversible Stirling and Ericsson heat punps. *J. Phys. D. Appl. Phys*., Vol. 35, pp 2668–2675.

[37] Tlili, I. (2012). Finite time thermodynamics evaluation of endoreversible Stirling heat engine at maximum power conditions. *Renew. Sust. Energy Rev*., Vol. 16, pp 2234–2241.

[38] Ladino-Luna, D., Portillo-Díaz, P., & Páez-Hernández, R.T. (2013). On the Efficiency for Non-Endoreversible Stirling and Ericsson Cycles. J. Modern Phys., Vol. 4, pp 1–7.

Chapter 5

Thermal Hysteresis Due to the Structural Phase Transitions in Magnetization for Core-Surface Nanoparticles

Rıza Erdem, Songül Özüm and Orhan Yalçın

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/61043

Abstract

One important issue raised in magnetism studies is the thermal response of various magnetic properties. This topic is known as the magnetic thermal hysteresis (MTH) which is principally associated with magnetic phase transitions. The MTH is of particular interest for both quantum and applied physics researches on magnetization of nanomaterials. Hysteresis of the temperature-induced structural phase transitions in some materials and nanostructures with first-order phase transitions reduces useful magnetocaloric effect to transform cycling between martensite (M) and austenite (A) phases under application. In additional, the size, surface and boundary effects on thermal hysteresis loops have been under consideration for the development of research on nanostructured materials. Experimental data indicate that nanostructured materials offer many interesting prospects for the magnetization data and for understanding of temperature-induced M-A phase transitions. In this chapter, we have presented a review of the the latest theoretical developments in the field of MTH related to the structural phase transitions for the core-surface nanoparticles based on the fundamental formulation of pair approximation in Kikuchi version.

Keywords: Thermal hysteresis, Nanoparticles, Martensite, Austenite, Pair approximation

1. Introduction

The phenomenon of hysteresis is encountered in many areas of physics. It is associated with the delay of the dynamic response of cooperative systems to external perturbation. During a heating–cooling process in a system, thermal hysteresis (TH) commonly appears accompanying phase transitions. In particular, it is regarded as a signature of the first-order phase transition. But, the TH is less known than the magnetic hysteresis (MH), which is another type

of hysteresis in magnetism [1]. Rao and Pandit [2] studied both TH and MH, and they found that TH seems to belong to a different universality class than MH in the same relaxational model.

One important issue raised in magnetism studies is the thermal response of various magnetic properties. This topic is known as the magnetic thermal hysteresis (MTH) or TH in magnetization, which is principally associated with magnetic phase transitions (MPTs). Many bulk materials with MPTs have been discovered to exhibit significant MTH behaviour. Among those materials, well known samples are spin crossover compounds [3–8], manganites [9], superconductors [10, 11], magnetic multilayers [12, 13], ferrimagnetic and metamagnetic alloys [14, 15], polycrystalline samples [16], manganite thin films [17] and manganite perovskites [18]. Thermal hysteresis occurs in these samples because the transitions between various magnetic phases occur at different temperatures for the heating and cooling processes. Furthermore, the temperature span of the TH can be substantially reduced by applying an external magnetic field. The thermoelastic austenite–martensite transformations in Heusler and shape memory alloys were also characterized by the TH loops [19–21].

Apart from the above investigations of bulk systems, the MTH is of particular interest for both quantum and applied physics researches on magnetization of nanomaterials. Several groups studied some important physical properties of various types of nanostructures using the properties of TH loops. For example, from the thermal response of conductivity for the gold nanoparticles (NPs), a phase transition phenomenon was revealed by a temperature criticality by Sarkar et al. [22]. Based on Ising-like treatment of the MTH and using Monte Carlo algorithm, Kawamoto and Abe [23] obtained smaller hysteresis width when the volume of the spin crossover NPs was decreased. Also, using the same treatment, the simulated hysteretic loops become closer to the experimental ones [24, 25].

On the other hand, hysteresis of the temperature-induced structural phase transitions in nanostructures with first-order phase transitions reduce useful magnetocaloric effect to transform cycling between martensite (M) and austenite (A) phases under application. In addition, the size, surface and boundary effects on thermal hysteresis loops have been under consideration for the development of research on nanostructured materials. Experimental data indicate that nanostructured materials offer many interesting prospects for the magnetization data and for understanding of temperature-induced martensite/austenite phase transitions.

In this chapter, we shall review the latest theoretical developments in the field of MTH related to the structural phase transitions for the core/surface (*C*/*S*) NPs based on the fundamental formulation of pair approximation in Kikuchi version.

2. Basics of the theoretical model

2.1. Definition of a nanoparticle with core/surface morphology

For a noninteracting spherical nanoparticle, arrays of spins are generally considered on a hexagonal lattice in 2D, as shown in Fig. 1 [26]. This structure may also be extended to hexagonal closed packed (hcp) lattice for any three-dimensional (3D) case which is not covered

http://dx.doi.org/10.5772/61043

in Fig. 1 but is illustrated in a recent publication by Yalçın et al. [27]. Similarly, a square lattice in 2D (shown in Fig. 2) is enlarged to simple cubic lattice (sc) in 3D for a cubic nanoparticle. For the number of shells in both structures, each lattice is related to the radius (R) of the nanoparticle [27–29]. Therefore, the value of R contains a number of shells and the size of a nanoparticle increases as the number of shells increases. The shells (R) and their numbers are only bounded to the nearest-neighbour pair exchange interactions (J) between spins. To provide the magnetization of the whole particle, each of the spin sites, which stand for the atomistic moments in the nanoparticle, are described by Ising spin variables that take on the values $S_i = \pm 1, 0$. For a core/surface (C/S) morphology, all spins in the nanoparticle are organized in three components that are core (C, filled circles), interface (or core-surface) (CS) and surface (S, empty circles) parts. The number of spins in these parts within the C/S-type nanoparticle are denoted by N_C, N_{CS} and N_S, respectively. But, the total number of spins (N) in a C/S nanoparticle covers only C and S spin numbers, i.e. $N = N_C + N_S$. On the other hand, the numbers of spin pairs for C, CS and S regions in 2D are defined by $N_P^C = (N_C\gamma_C/2) - N_{CS}$, $N_P^{CS} = 2N_{CS}\gamma_{CS}/2$ and $N_P^S = N_S\gamma_S/2$, respectively, where the lattice coordination numbers of the regions are $\gamma_C = 6$, $\gamma_{CS} = \gamma_S = 2$ for hexagonal lattice and $\gamma_C = 4$, $\gamma_{CS} = 2$, $\gamma_S = 0$ for square lattice.

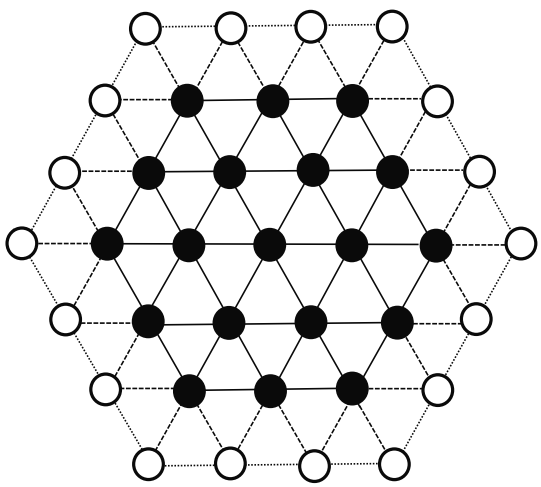

Figure 1. Schematic representation of a nanoparticle on a hexagonal lattice in 2D exhibiting three shells of spins. Filled and empty circles represent the core and surface atoms while solid, dashed and dotted lines correspond to core, core/surface and surface pairs, respectively.

2.2. Blume–Emery–Griffiths model

The Blume–Emery–Griffiths (BEG) model is one of the well-known spin lattice models in equilibrium statistical mechanics. It was originally introduced with the aim to account for phase separation in helium mixtures [30]. Besides various thermodynamic properties, the model has been extended to study the structural phase transitions in many bulk systems. By

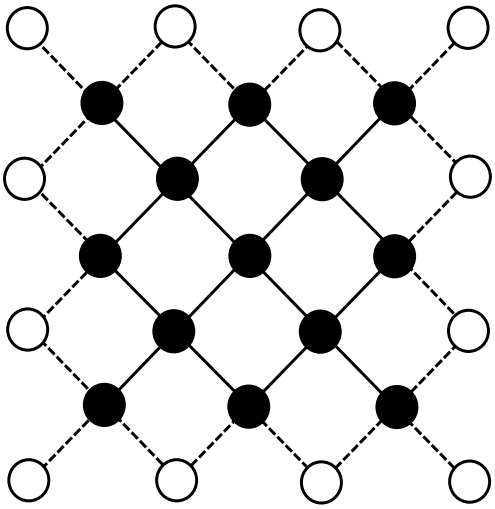

Figure 2. Same as Fig. 1 but for a nanoparticle on a square lattice.

means of mean-field theory (MFT) and Monte Carlo (MC) simulations, magnetostructural phase transitions in some alloys were described via degenerate BEG models in terms of magnetoelastic interactions [31]. In the light of above applications, we have recently used the ordinary BEG model for the investigation of MT/AT transitions in NP systems and observed the behaviours of the MTH loops [28, 29]. In the following, we mention briefly the definition of the BEG Hamiltonian and show clearly how it is modified for the *C*/*S* NPs with hexagonal and square lattice structures.

For a spin configuration $\{S_i\}$, the ordinary BEG model is described by the Hamiltonian

$$H\{S_i\} = -J\sum_{\langle i,j\rangle} S_i S_j - K\sum_{\langle i,j\rangle} S_i^2 S_j^2 - D\sum_{\langle i,j\rangle}(S_i^2 + S_j^2) - h\sum_{\langle i,j\rangle}(S_i + S_j), \tag{1}$$

where $<i, j>$ indicates a sum over the nearest neighbours. J, K, D, h denote, respectively, the dipolar (or bilinear) exchange energy between nearest-neighbour spins, the quadrupolar (or biquadratic) exchange coupling, the single-ion anisotropy constant (or crystal-field parameter) and the external magnetic field. The above parameters are in units of kT (k Boltzmann constant and T temperature). Using various techniques, the model has been analyzed globally, for both negative (J, $K < 0$) and positive (J, $K > 0$) interactions, to obtain the equilibrium phase properties [32, 33]. Here, $J < 0$ and $J > 0$ cases correspond to ferromagnetic (FM) and antiferromagnetic (AFM) dipolar interactions, respectively. When $J = 0$, a paramagnetic (PM) character exists in the system for all temperatures. Similarly, the cases $K < 0$ and $K > 0$ describe the repulsive and attractive biquadratic interactions, respectively, and determine the rich phase diagrams with multicritical topology [33]. In the case of $K = 0$, the model is known as the Blume–Capel (BC) model.

http://dx.doi.org/10.5772/61043

The model Hamiltonian (Eq. 1) is now divided into three parts for the *C*/*S* NPs given by $H = H_C + H_{CS} + H_S$ with

$$\begin{aligned} H_C &= -J_C \sum_{\langle i,j \rangle} S_i S_j - K_C \sum_{\langle i,j \rangle} S_i^2 S_j^2 - D_C \sum_{\langle i,j \rangle} (S_i^2 + S_j^2) - h \sum_{\langle i,j \rangle} (S_i + S_j), \\ H_{CS} &= -J_{CS} \sum_{\langle i,j \rangle} S_i \sigma_j - K_{CS} \sum_{\langle i,j \rangle} S_i^2 \sigma_j^2, \\ H_S &= -J_S \sum_{\langle i,j \rangle} \sigma_i \sigma_j - K_S \sum_{\langle i,j \rangle} \sigma_i^2 \sigma_j^2 - D_S \sum_{\langle i,j \rangle} (\sigma_i^2 + \sigma_j^2) - h \sum_{\langle i,j \rangle} (\sigma_i + \sigma_j), \end{aligned} \tag{2}$$

where S_i and σ_i are named as the *C* and *S* spin variables, respectively. In Eq. (2) J_C, J_{CS}, J_S are the bilinear and K_C, K_{CS}, K_S are the biquadratic exchange interactions for *C, CS* and *S* spins, respectively. Moreover, single-ion anisotropy parameters for *C* and *S* spins are denoted by the letters D_C and D_S, respectively. If $J_C = J_{CS} = J_S = J_0$ and $K_C = K_{CS} = K_S = K_0$ the particle is known as homogeneous (HM) nanoparticle while one of the conditions $J_C \neq J_{CS} \neq J_S$, $J_C = J_{CS} \neq J_S$, $J_C \neq J_{CS} = J_S$, $J_{CS} \neq J_C = J_S$, $K_C \neq K_{CS} \neq K_S$, $K_C = K_{CS} \neq K_S$, $K_C \neq K_{CS} = K_S$, $K_{CS} \neq K_C = K_S$ corresponds to a composite (CM) nanoparticle.

2.3. Fundamental formulation of pair approximation in Kikuchi version

In the pair approximation proposed by Kikuchi [34], each spin case is indicated by p_i, which is also so entitled the point or state variables. These point/state variables obey the normalization relation

$$\sum_i p_i = 1. \tag{3}$$

Using the pair correlations between spins, another types of internal (or bond/pair) variables denoted by P_{ij} are introduced. P_{ij} (with a symmetry $P_{ij} = P_{ji}$) means the medial number of the states in which the first members of the nearest-neighbour pair is in state *i* and second member in state *j*. The bond variables are also normalized by

$$\sum_{i,j} P_{ij} = 1, \tag{4}$$

and connected with state variables through the relations

$$p_i = \sum_{i,j} P_{ij}. \tag{5}$$

In order to determine an expression for the bond variables, we define the interaction energy (E) and entropy (S_E) of system in terms of these variables as

$$\beta E = N\frac{\gamma}{2}\sum_{i,j}\varepsilon_{ij}P_{ij}, \tag{6}$$

$$S_E = Nk\left((\gamma-1)\sum_{i,j}p_i\ln(p_i) - \frac{\gamma}{2}\sum_{i,j}P_{ij}\ln(P_{ij})\right), \tag{7}$$

where $\beta=1/kT$, γ is the coordination number for a lattice site and N is the number of these sites. The ε_{ij} parameters in Eq. (6) are called the bond energies for the spin pairs (i, j) and determined from Eq. (1). The free energy (Φ) per site can be found from

$$\Phi = \frac{\beta F}{N} = \frac{\beta}{N}(E - TS_E). \tag{8}$$

The minimization of Eq. (8) with respect to P_{ij} ($\partial\Phi/\partial P_{ij}=0$) leads to the following set of self-consistent equations for the system at equilibrium:

$$P_{ij} = Z^{-1}(p_i p_j)^{(\gamma-1)/\gamma}\exp(-\beta\varepsilon_{ij}) \equiv Z^{-1}e_{ij}, \tag{9}$$

where Z is the partition function:

$$Z = \exp(2\beta\lambda/\gamma) = \sum_{i,j}e_{ij}. \tag{10}$$

Here, λ is introduced to maintain the normalization condition. In many works in the literature, Eq. (9) was easily applied to investigate magnetic properties of various spin models for the bulk materials [35, 36]. But, for the applications to the magnetic NPs systems, one needs to define the energy parameters ε_{ij} appearing in Eq. (9) as

$$\varepsilon_{ij} = N_P^C\varepsilon_{ij}^C + N_P^{CS}\varepsilon_{ij}^{CS} + N_P^S\varepsilon_{ij}^S, \tag{11}$$

where bond energies ε_{ij}^C, ε_{ij}^{CS} and ε_{ij}^S of three regions are found using Eq. (2) as listed in Table 1. The average magnetization (M) of the nanoparticle is the excess of one orientation over the other orientation, also named the dipole moment. It is found from the definition

http://dx.doi.org/10.5772/61043

$$M = P_{++} + P_{+0} + P_{+-} - (P_{-+} + P_{-0} + P_{--}) \tag{12}$$

Using the numerical solutions of Eq. (9) by the iteration technique, the MTH curves are drawn easily from Eq. (12). For some selected exchange energies, the MTH behaviours and the temperature-induced M–A phase transitions within the *C*/*S* smart NPs are reproduced from our recent publications [28, 29] as in Figs. 3–6.

	C	**CS**	**S**
ε_{++}	$-J_C - K_C - 2D_C - 2h$	$-J_{CS} - K_{CS}$	$-J_S - K_S - 2D_S - 2h$
ε_{+0}	$-D_C - h$	0	$-D_S - h$
ε_{+-}	$+J_C - K_C - 2D_C$	$+J_{CS} - K_{CS}$	$+J_S - K_S - 2D_S$
ε_{0+}	$-D_C - h$	0	$-D_S - h$
ε_{00}	0	0	0
ε_{0-}	$-D_C + h$	0	$-D_S + h$
ε_{-+}	$+J_C - K_C - 2D_C$	$+J_{CS} - K_{CS}$	$+J_S - K_S - 2D_S$
ε_{-0}	$-D_C + h$	0	$-D_S + h$
ε_{--}	$-J_C - K_C - 2D_C + 2h$	$-J_{CS} - K_{CS}$	$-J_S - K_S - 2D_S + 2h$

Table 1. Bond energies for the *C*/*S* nanoparticle

3. Calculations and discussion

3.1. Thermal hysteresis for hexagonal nanoparticles

We firstly analyze the MTH loops for the homogeneous hexagonal nanoparticles (HM-HNPs) using $J_C = J_{CS} = J_S = J_0 = 1.0$; $K_C = K_{CS} = K_S = K_0 = -0.6$ under zero magnetic field ($h = 0.0$). Fig. 3 represents the behaviours of these loops for the particles with $R = 4$ and $R = 5$ shells. In the figure, each coloured curve is drawn for one value of single-ion anisotropy (D_0) with $D_C = D_S = D_0$. The thermal behaviour of the particles' magnetization is somewhat different from that of bulk materials. We see clearly (Figs. 3a and 3b) that the magnetization returns to the original point after a complete cycle. The cycles are a specific class of both major and minor hysteretic loops, for which the temperature is reversed only once. For a structural transition, the heating/cooling modes are to be distinguished. Each coloured curve represents both heating and cooling processes according to directions of the arrows. Therefore, the calculations start at a low temperature below the austenite start temperature (A_S) where the magnetization starts to increase sharply until the austenite finish temperature (A_F) is reached. The structure between these two temperatures with $A_S < A_F$ corresponds to a strongly austenite phase. After passing through a maximum, the magnetization decreases as the temperature increases and converges

to zero from another austenite finish temperature A_{F1} to austenite start temperature A_{S1} where we observe a weakly austenite phase with $A_{F1} < A_{S1}$. On the other hand, the cooling process concerns the structural changes associated with martensitic transitions. This may cause an abrupt change of magnetization where a weak martensite phase occurs. The new characteristic temperatures related to the increase in magnetization are denoted by M_{S1} and M_{F1} ($M_{S1} < M_{F1}$) corresponding to martensitic start and martensitic finish temperatures, respectively. As is illustrated at the right columns of each figure in Fig. 3, each coloured curve for the heating/cooling processes coincides with each other at M_{S1} and A_{F1} while they do not coincide at the temperatures M_{F1} and A_{S1}. Further decreasing of the temperature leads to a strongly martensite state, which starts at M_S and ends at M_F. Thus, we have two separate MTH loops; one is at the lower temperature (or the major thermal hysteresis) while the other occurs at larger temperatures (minor thermal loops). For the major case, we note that the loops shift to the higher temperatures and the actual loop area becomes smaller after the decrease of D_0. But, the minor loop shifts to the lower temperatures while the amount of the hysteresis remains approximately the same. A direct comparison of the left/right panels of Fig. 3a with those of Fig. 3b shows how the particle size correlates with all characteristic temperatures.

For understanding the features associated with the MTH behaviours in composite hexagonal nanoparticles (CM-HNPs), we have proceeded to calculate the magnetization as a function of temperature for the same system, but using $J_{CS} = -J_0$ instead of $J_{CS} = J_0$, whose results are shown in Figs. 4a and 4b. Again, we consider four- and five-shell NPs under zero magnetic field (h = 0.0). The new values of D_0 associated with the observation of MT/AT structural phase transitions are listed in each figure. In this case, the small positive values of the single-ion anisotropy are needed for the generic property of MTH loops presented in the previous figure. As demonstrated by Fig. 4, the AFM *C*/*S* dipolar interaction with $J_{CS} = -J_0$ also influences the heights, widths and the positions of the major and minor loops presented in Fig. 3. This fact states that, at constant R and K_0, all characteristic temperatures are decreasing functions of D_0 which determine the smoothness of the reversal. The small hysteresis loops seen in Fig. 4a indicate that the friction to resist the structural transformation is small. According to Hu et al. [15], the small thermal hysteresis is the aspiration of an engineer in applying the materials in magnetocaloric refrigeration. This makes the present CM-HNPs attractive for real applications in new technologies [37].

3.2. Thermal hysteresis for cubic nanoparticles

Another class of small particles known in the literature are called cubic nanoparticles (CNPs). The CNPs attach face-to-face to the surrounding particles. They can form a 2D square array instead of hexagonal ones (Fig. 2). The square array is one of the closed-packed arrangements and the CNPs with a square array have large surface to volume ratios. Owing to their different electronic properties, in recent years, the CNPs received much attention for applications mostly in material science, sensor technology and semiconductor devices. But, the magnetic properties have been very sensitive to the particle shape due to dominating role of surface anisotropy in its magnetization. Compared to the spherical nanoparticles (SNPs), the flat surface of CNPs enabled the surface metal captions to possess a more symmetric coordination. So, the surface

http://dx.doi.org/10.5772/61043

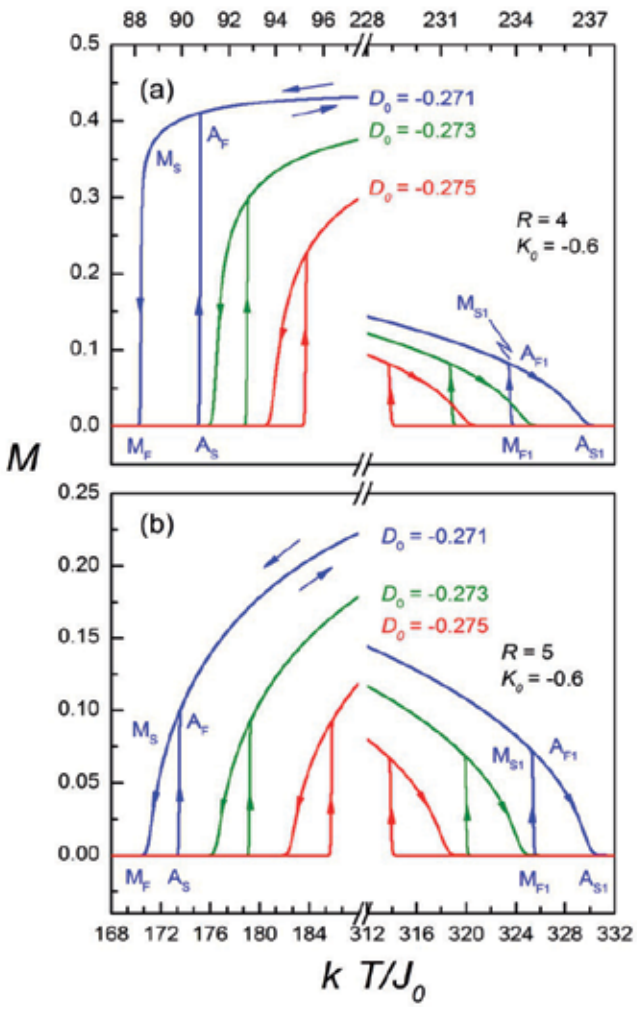

Figure 3. Magnetic thermal hysteresis loops for the HM-HNPs with (a) $R = 4$ and (b) $R = 5$ shells. $J_C = J_{CS} = J_S = J_0 = 1.0$; $K_C = K_{CS} = K_S = K_0 = -0.6$; $D_C = D_S = D_0$ and $h = 0.0$.

anisotropy in the CNPs should be much smaller than the one in HNPs and SNPs. If the magnetic anisotropy of the CNPs is cubic, all such six directions are magnetically identical, and hence the magnetically ordered assembly is greatly simplified. The anisotropy of the growth rate can be ascribed to a different adhesion of the stabilizer on the growing surface. The stabilizer species is the only parameter which was changed to obtain cubical shapes instead of spheres. Because of the above properties, the single-domain CNPs with surface anisotropy were widely investigated using various numerical methods [38–42].

For the sake of comparison with that of HNPs, we here reestablish the MTH curves of the homogeneous cubic nanoparticles (HM-CNPs) and composite cubic nanoparticles (CM-CNPs) by using the same procedure as in the preceding section (Figs. 3 and 4) and focus on the

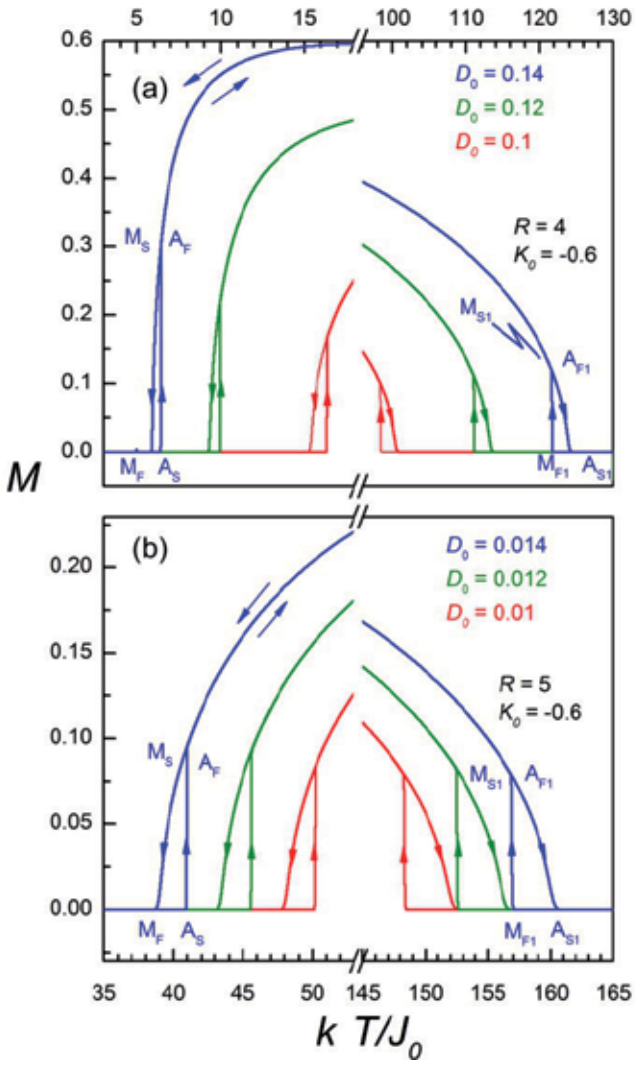

Figure 4. Same as Fig. 3 but for the CM-HNPs with $J_{CS} = -J_0$.

important difference between the martensitic and austenitic transition temperatures, derived from these hysteresis loops, for the CNPs and HNPs. The general aspects of the MTH loops for the CNPs displayed in Figs. 5 and 6 are similar to one for the HNPs given in Figs. 3 and 4. But, the loops are calculated using bigger particles ($R = 6$ and $R = 7$ shells) with greater single-ion anisotropy values, which determine the smoothness of the reversal. It is observed that the transition temperatures diminish when reducing the nanoparticles size in correlation with a smoother conversion, seen in Figs. 5a and 5b. In addition, while the NP sizes are increasing, the single-ion anisotropy parameter also increases and thermal hysteresis loop appears as a rectangular-shaped loop ($D_0 = -0.265$) as shown in Fig. 5b. The single-ion anisotropy parameter

http://dx.doi.org/10.5772/61043

decreases at the larger radius values for the CM-CNPs in Figs. 6a and 6b. If we compare the HM-NPs and CM-NPs, it can be seen that magnetization shows almost the same values but, the transition temperatures of MTH loops for HM-NPs occurred at higher values.

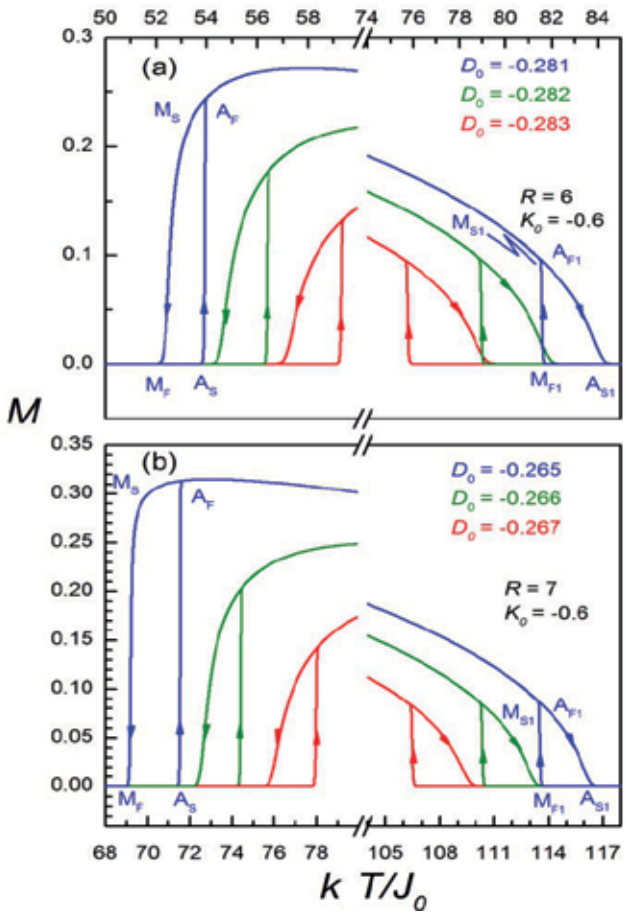

Figure 5. Same as Fig. 3 but for a HM-CNPs with (a) $R = 6$ and (b) $R = 7$ shells.

In general, the thermal hysteresis becomes weaker but, nevertheless, does not disappear completely with increasing NP sizes for the CM-NPs. All loops widen as R decreases for both HM-HNPs and HM-CNPs. Also, a large decrease in transition temperatures occur when the particle structure changes from hexagonal to cubic array. The CNPs ensure that all magnetic properties are well understood because of their easily recognizable magnetization axes and well-defined crystallographic surfaces.

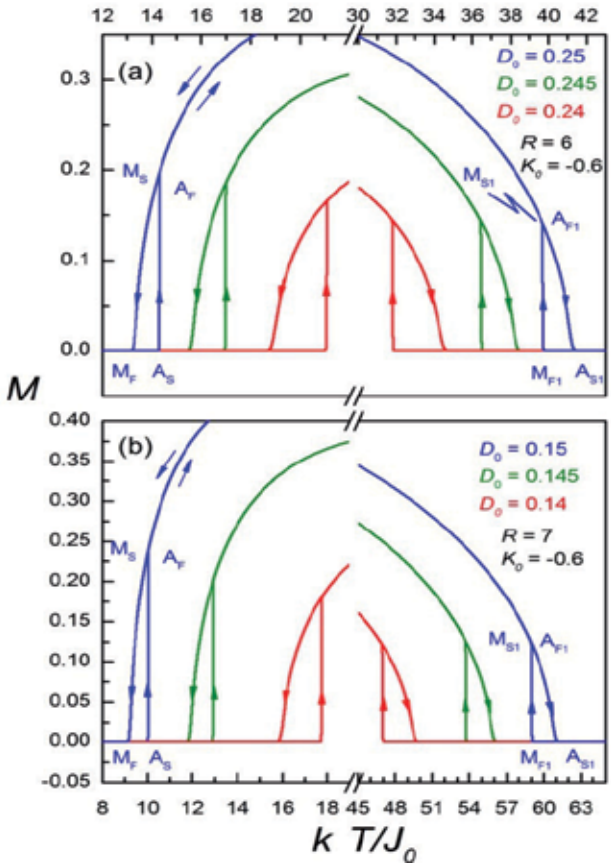

Figure 6. Same as Fig. 4 but for a CM-CNPs with (a) $R = 6$ and (b) $R = 7$ shells.

4. Conclusion

In this chapter, we have presented a review of thermal reversal properties of the *C*/*S* NPs related with the martensitic/austenitic phase transitions. These properties have been recently studied using a powerful statistical mechanics approach called the pair approximation technique. According to the theoretical calculations performed for the hexagonal and cubic NPs at a given particle radius, magnetization versus temperature variations are essentially two closed loops, i.e. MTH curves or loops because of the structural phase transitions from M state to A state (or from A phase to M phase) during heating–cooling processes.

Our review draws a number of important physical properties for the MTH curves regarding the sign of the quadrupolar interactions (K_0) and the values of sing-ion anisotropy (D_0) within the NPs. (1) During heating–cooling processes, we observe TH behaviours for the particles' magnetization, which is indicative of a first-order transition when the quadrupolar interaction is repulsive. The positions and shapes of the loops and the loop area depend on the values of crystal-field parameter as well as the size of the particle. (2) We also show that the thermal

evolution of the magnetization separates the TH loops into two different regions, which have different ordering temperatures. In other words, we observe two TH loops; one (major loop) is at lower temperatures and the other (minor loop) at higher temperatures. (3) We fully explain the thermal reversal process and show that the temperatures at which the NPs reverses its magnetization depend on the applied external magnetic fields while the single-ion anisotropy determines the width and smoothness of the reversal. (4) Most important is the fact that, by varying *C*/*S* dipolar interaction from FM-type to AFM-type as well as by changing crystal field parameter from negative to positive values, a very small TH loop can be obtained and the structural transition temperatures can be turned over a small range of working temperatures.

Acknowledgements

One of the authors (RE) acknowledges the financial support from the Scientific Research Projects Coordination Unit of Akdeniz University.

Author details

Rıza Erdem[1*], Songül Özüm[2] and Orhan Yalçın[3]

*Address all correspondence to: rerdem@akdeniz.edu.tr

1 Department of Physics, Akdeniz University, Antalya, Turkey

2 Institute of Sciences, Niğde University, Niğde, Turkey

3 Department of Physics, Niğde University, Niğde, Turkey

References

[1] Mamiya H, Jeyadevan B. Magnetic hysteresis loop in a superparamagnetic state. IEEE Transactions on Magnetics. 2014; 50: 4001604. DOI: 10.1109/TMAG.2013.2274072.

[2] Rao M, Pandit R. Magnetic and thermal hysteresis in the *O(N)*-symmetric $(\Phi)^2$ model. Physical Review B. 1991; 43; 3373–3386. DOI: 10.1103/PhysRevB.43.3373.

[3] Enachescu C, Machado HC, Menendez N, Codjovi E, Linares J, Varret F, Stancu A. Static and light induced hysteresis in spin-crossover compounds: experimental data and application of Preisach-type models. Physica B. 2001; 306: 155–160. DOI: 10.1016/S0921-4526(01)00996-6.

[4] Linares J, Enachescu C, Boukheddaden K, Varret F. Monte Carlo entropic sampling applied to spin crossover solids: the squareness of the thermal hysteresis loop. Polyhedron. 2003; 22: 2453–2456. DOI: 10.1016/S0277-5387(03)00219-5.

[5] Tanasa R, Enachescu C, Stancu A, Linares J, Codjovi E, Varret F, Haasnoot J. First-order reversal curve analysis of spin-transition thermal hysteresis in terms of physical-parameter distributions and their correlations. Physical Review B. 2005; 71: 014431(1)-014431(9). DOI: 10.1103/PhysRevB.71.014431.

[6] Enachescu C, Tanasa R, Stancu A, Varret F, Linares J, Codjovi E. First-order reversal curves analysis of rate-dependent hysteresis: The example of light-induced thermal hysteresis in a spin-crossover solid. Physical Review B. 2005; 72: 054413(1)–054413(7). DOI: 10.1103/PhysRevB.72.054413.

[7] Stoleriu L, Chakraborty P, Hauser A, Stancu A, Enachescu C. Thermal hysteresis in spin-crossover compounds studied within the mechanoelastic model and its potential application to nanoparticles. Physical Review B. 2011; 84: 134102(1)–134102(9). DOI: 10.1103/PhysRevB.84.134102.

[8] Rotaru A, Graur A, Rotaru G–M, Linares J, Garcia Y. Influence intermolecular interactions and size effect on LITH-FORC diagram in 1D spin-crossover compounds. Journal of Optoelectronics and Advanced Materials. 2012; 14: 529–536.

[9] Dho J, Kim WS, Hur NH. Anomalous thermal hyteresis in magnetization and resistivity of $La_{1-x}Sr_xMnO_3$. Physical Review Letters. 2001; 87: 187201(1)–187201(4). DOI: 10.1103/PhysRevLett.87.187201.

[10] Panagopoulos C, Majoros M, Petrović AP. Thermal hyteresis in the normal-state magnetization of $La_{2-x}Sr_xCuO_4$. Physical Review B. 2004; 69: 144508(1)–144508(7). DOI: 10.1103/PhysRevB.69.144508.

[11] Hissa J, Kallio A. Comment on ''Thermal hyteresis in the normal-state magnetization of $La_{2-x}Sr_xCuO_4$''. Physical Review B. 2005; 72: 176501(1)–176501(3). DOI: 10.1103/PhysRevB.72.176501.

[12] Camley RE, Lohstroh W, Felcher GP, Hosoito N, Hashizume H. Tunable thermal hysteresis in magnetic multilayers: magnetic superheating and supercooling. Journal of Magnetism and Magnetic Materials. 2005; 286: 65–71. DOI: 10.1016/j.jmmm.2004.09.041.

[13] Dantas AL, Camley RE, Carriço AS. Magnetic thermal hysteresis in $Fe_m/Dy_n/Fe_m$ and $Gd_m/Dy_n/Gd_m$ trilayers. Physical Review B. 2007; 75: 094436(1)–094436(6). DOI: 10.1103/PhysRevB.75.094436.

[14] Andrés JP, González JA, Hase TPA, Tanner BK, Riveiro JM. Artificial ferrimagnetic structure and thermal hysteresis in $Gd_{0.47}/Co_{0.53}/Co$ multilayers. Physical Review B. 2008; 77: 144407(1)–144407(7). DOI: 10.1103/PhysRevB.77.144407.

[15] Hu FX, Wang J, Shen J, Gao B, Sun JR, Shen BG. Large magnetic entropy change with small thermal hyteresis near room temperature in metamagnetic alloys $Ni_{51}Co_{49-x}In_x$. Journal of Applied Physics. 2009; 105: 07A940(1)–07A940(3). DOI: 10.1063/1.3073951.

[16] Trung NT, Ou ZQ, Gortenmulder TJ, Tegus O, Buschow KHJ, Brück E. Tunable thermal hysteresis in MnFe(P,Ge) compounds. Journal of Applied Physics. 2009; 94: 102513(1)–102513(3). DOI: 10.1063/1.3095597.

[17] Singh S, Fitzsimmons MR, Lookman T, Thompson JD, Jeen H, Biswas A, Roldan MA, Varela M. Magnetic nonuniformity and thermal hyteresis of magnetism in a manganite thin film. Physical Review Letters. 2012; 108: 077207(1)–077207(5). DOI: 10.1103/PhysRevLett.108.077207.

[18] Wang L, Cui YG, Wan JF, Zhang JH, Rong YH. Magnetic thermal hysteresis due to paramagnetic–antiferromagnetic transition in Fe-24.4Mn-5.9Si-5.1Cr alloy. AIP Advances. 2013; 3: 082126(1)–082126(8). DOI: 10.1063/1.4819483.

[19] Krenke T, Acet M, Wassermann EF, Moya X, Moñosa L, Planes A. Martensitic transitions and the nature of ferromagnetism in the austenitic and martensitic states of Ni-Mn-Sn alloys. Physical Review B. 2005; 72: 014412(1)–014412(9). DOI: 10.1103/PhysRevB.72.014412.

[20] Shamberger PJ, Ohuchi FS. Hysteresis of the martensitic phase transition in magnetocaloric effect Ni-Mn-Sn alloys. Physical Review B. 2009; 79: 144407(1)–144407(9). DOI: 10.1103/PhysRevB.79.144407.

[21] Lakhani A, Dash S, Banerjee A, Chaddah P, Chen X, Ramanujan RV. Tuning the austenite and martensite phase fraction in ferromagnetic shape memory alloy ribbons of $Ni_{45}Co_5Mn_{38}Sn_{12}$. Applied Physics Letters. 2011; 99: 242503(1)–242503(3). DOI: 10.1063/1.3669510.

[22] Sarkar T, Roy S, Bhattacharya J, Bhattacharya D, Mitra CK, Dasgupta AKr. Thermal hysteresis of some important physical properties of nanoparticles. Journal of Colloid and Interface Science. 2008; 327: 224–232. DOI: 10.1016/j.jcis.2008.07.050.

[23] Kawamoto T, Abe S. Thermal hysteresis loop of the spin-state in nanoparticles of transition metal complexes: Monte Carlo simulations on an Ising-like model. Chemical Communications. 2005; No. 31, 3933–3935. DOI: 10.1039/b506643c.

[24] Atitoaie A, Tanasa R, Enachescu C. Size dependent thermal hysteresis in spin crossover nanoparticles reflected within a Monte-Carlo based Ising-like model. Journal of Magnetism and Magnetic Materials. 2012; 324: 1596–1600. DOI: 10.1016/j.jmmm.2011.12.011.

[25] Atitoaie A, Tanasa R, Stancu A, Enachescu C. Study of spin crossover nanoparticles thermal hysteresis using FORC diagrams on an Ising-like model. Journal of Magnetism and Magnetic Materials. 2014; 368: 12–18. DOI: 10.1016/j.jmmm.2014.04.054.

[26] Rego LGC, Figueiredo W. Magnetic properties of nanoparticles in the Bethe–Peierls approximation. Physical Review B. 2001; 64: 144424(1)–144424(7). DOI: 10.1103/PhysRevB.64.144424.

[27] Yalçın O, Erdem R, Demir Z. Magnetic properties and size effects of spin-1/2 and spin-1 models of core-surface nanoparticles in different type lattices. In: Abbass Hashim, editor. Smart Nanoparticles Technology: InTech; 2012. p. 541–560. DOI: 10.5772/34706.

[28] Yalçın O, Erdem R, Özüm S. Origin of the martensitic and austenitic phase transition in core-surface smart nanoparticles with size effects and hysteretic splitting. Journal of Applied Physics. 2014; 115: 054316(1)–054316(7). DOI: 10.1063/1.4864489.

[29] Özüm S, Yalçın O, Erdem R, Bayrakdar H, Eker HN. Martensitic and austenitic phase transformations in core-surface cubic nanoparticles. Journal of Magnetism and Magnetic Materials. 2015; 373: 217–221. DOI: 10.1016/j.jmmm.2014.03.044.

[30] Blume M, Emery VJ, Griffiths RB. Ising model for the λ transition and phase separation in He^3-He^4 mixtures. Physical Review A. 1971; 4: 1071–1077. DOI: 10.1103/PhysRevA.4.1071.

[31] Castán T, Vives E, Lindgård PA. Modeling premartensitic effects in Ni_2MnGa: A mean-field and Monte Carlo simulation study. Physical Review B. 1999; 60: 7071–7084. DOI: 10.1103/PhysRevB.60.7071.

[32] Berker AN, Wortis M. Blume–Emery–Griffiths–Potts model in two dimensions: phase diagram and critical properties from a position-space renormalization group. Physical Review B. 1976; 14: 4946–4963. DOI: 10.1103/PhysRevB.14.4946.

[33] Hoston W, Berker AN. Multicritical phase diagrams of the Blume–Emery–Griffiths model with repulsive biquadratic coupling. Physical Review Letters. 1991; 67: 1027–1030. DOI: 10.1103/PhysRevLett.67.1027.

[34] Kikuchi R. A theory of cooperative phenomena. Physical Review. 1951; 81. 988–1003.

[35] Meijer PHE, Keskin M, Bodegom E. A simple model for the dynamics towards metastable states. Journal of Statistical Physics. 1986; 45: 215–232. DOI: 10.1007/BF01033088.

[36] Erdinç A, Keskin M. Equilibrium and nonequilibrium behavior of the spin-1 Ising model in the quadrupolar phase. Physica A. 2002; 307: 453–468. DOI: 10.1016/S0378-4371(01)00620-3.

[37] Brück E, Trung NT, Ou ZQ, Buschow KHJ. Enhanced magnetocaloric effects and tunable thermal hysteresis in transition metal pnictides. Scripta Materialia. 2012; 67: 590–593. DOI: 10.1016/j.scriptamat.2012.04.037.

[38] Yanes R, Chubykalo-Fesenko O, Kachkachi H, Garanin DA, Evans R, Chantrell RW. Effective anisotropies and energy barriers of magnetic nanoparticles with Néel sur-

face anisotropy. Physical Review B. 2007; 76: 064416(1)–064416(13). DOI: 10.1103/PhysRevB.76.064416.

[39] Kita E, Oda T, Kayano T, Sato S, Minagawa M, Yanagihara H, Kishimoto M, Mitsumata C, Hashimoto S, Yamada K, Ohkohchi N. Ferromagnetic nanoparticles for magnetic hyperthermia and thermoablation therapy. Journal of Physics D: Applied Physics. 2010; 43: 474011(1)–474011(9). DOI: 10.1088/0022-3727/43/47/474011.

[40] Knittel A, Franchin M, Fischbacher T, Nasirpouri F, Bending SJ, Fangohr H. Micromagnetic studies of three-dimensional pyramidal shell structures. New Journal of Physics. 2010; 12: 113048(1)–113048(23). DOI: 10.1088/1367-2630/12/11/113048.

[41] Carrey J, Mehdaoui B, Respaud M. Simple models for dynamic hysteresis loop calculations of magnetic single-domain nanoparticles: application to magnetic hyperthermia optimization. Journal of Applied Physics. 2011; 109: 083921(1)–083921(17). DOI: 10.1063/1.3551582.

[42] Lee CF, Chang CL, Yang JC, Lai HY, Chen CH. Morphological determination of face-centered-cubic metallic nanoparticles by X-ray diffraction. Journal of Colloid and Interface Science. 2012; 369: 129–133. DOI: 10.1016/j.jcis.2011.12.053.

Chapter 6

Information Thermodynamics and Halting Problem

Bohdan Hejna

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/61900

Abstract

The formulations of the *undecidability* of the Halting Problem assume that the computing process, being observed, the description of which is given on the input of the 'observing' Turing Machine, is, at any given moment, the exact copy of the computing process running in the observing machine itself (the Cantor diagonal argument). In this way an *infinite cycle* is created shielding what is to be possibly discovered - the possible infinite cycle in the observed computing process. By this type of our consideration and in the thermodynamic sense the equilibrium status of a certain thermodynamic system is described or, even created. This is a thermodynamic image of the Cantor diagonal method used for seeking a possible infinite cycle and which, as such, has the property of the Perpetuum Mobile - the structure of which is recognizable and therefore we can avoid it. Thus we can show that it is possible to recognize the infinite cycle as a certain *original* equilibrium, but with a *'step-aside'* or a time delay in evaluating the trace of the observed computing process.

The trace is a record of the se*quence of configurations* of the observed Turing machine. These configurations can be simplified to their *common configuration types*, creating now a word of a *regular language*. Furthermore, the control unit of any Turing Machine is a *finite automaton*. Both these facts enable the *Pumping Lemma* in the observing Turing Machine to be usable. In compliance with the Pumping Lemma, we know (the observing Turing Machine knows) that certain common configuration types must be *periodically repeated* in the case of the *infinite length* of their regular language. This fact enables (in a finite time) us (the observing Turing Machine) to decide that the observed computing process has entered into an infinite cycle.

Considerations of the real sense of the Gibbs Paradox are used to illustrate the idea of the term 'step-aside' which is our main methodological tool for looking for the infinite cycle in a Turing computing process and which enables us to avoid the commonly used attempts to solve the Halting Problem.

Keywords: Heat and Information Entropy, Observation, Carnot Cycle, Information Channel, Turing Machine, Infinite Cycle

1. Introduction

The formulations of the *undecidability* of the Halting Problem assume that the computing process, being observed, the description of which is given on the input of the 'observing' Turing Machine, is the exact copy of the computing process running in the observing machine itself (the Cantor diagonal argument in the Minski's proof [18]). By this way the *Auto-Reference* or an *infinite cycle* in computing sense or the *Self-Observation* in information sense or an analogue of the *stationary (equilibrium) state* in thermodynamic sense is created, *shielding* now what is to be possibly discovered - the infinite cycle in the observed computing process - also for its normal input. This shield is the real result of the Cantor diagonal argument which has been used for solving the Halting Problem, but on the contrary, creates it [18]. This shield is, also, a certain image of the sought possible infinite cycle. This shield could be, in the thermodynamic point of view, ceased or ended whether the performance of the Perpetuum Mobile functionality was possible, which is not (when, e.g., the equation $x = x + 1$ would be solvable) or by an external activity or approach.

This situation is recognizable and as such is *decidable* and solvable in all cases of its realizations by a certain *'step-aside'*. For this, we use the previously studied [5, 6, 9, 11] congruence between a cyclical thermodynamic process represented by the Carnot Cycle and a repeatable information transfer represented by the Shannon Transfer Chain but we enrich this effort now by their another congruence with the computing process running in the Turing Machine. Considerations of Gibbs Paradox [7, 8, 11] are used to illustrate the main idea of the term 'step-aside' which is our main methodological tool for looking for an infinite cycle in a Turing computing process and which enables us to avoid the traditional attempts of solving the Halting Problem. The gap in their formulations is due to that fact that they assume that the computing process observes itself by following itself in the same 'time-click' of its activity. For this case the claim of the unsolvability of such a situation is, of course, valid. But with a *time delay* or the 'step-aside' [or a memory (a storage)] considered it works with 'another' data and 'is able to see' on its own previous activity as the 'normal' input data. The idea of the 'step-aside' is based just on the result giving the solution to Gibbs Paradox and its information meaning [7, 8, 11, 12, 14], and enables us to be in compliance with the *II. Principle of Thermodynamics* during such a process.

Thus, we show that it is possible to recognize the infinite cycle, but with the *time delay* or a *staging* (instead of the time delay the *staging* is usable, lastng a longer time interval in each of its repetition) in evaluating the *trace* of the observed computing process. The trace is a message or a record, both about the input data and about the structure of the computing process being observed (the *listing*, the *cross-references* and the *memory dump* in the language of programmers). In this phase the *observing* Turing Machine (we ourselves) is raising the question: "Is there an infinite cycle?" Following the trace the observing machine gains the answer. In our case, the trace is a *recorded sequence of configurations* of the *observed* Turing Machine. These configurations can be simplified to their *general configuration types* which create now a word of a *regular language* [12, 14]. Furthermore, the control unit of any Turing Machine is a *finite automaton*. Both these facts enable the *Pumping Lemma* in the observing Turing Machine to be used. In

compliance with the Pumping Lemma, we know (the observing Turing Machine knows) that certain general configuration types must be *periodically repeated* in the case of the *infinite length* of their regular language. This fact enables us (the observing Turing Machine) to decide that the observed computing process has entered into an infinite cycle. This event is performed in a finite time and is, by this way, recognizable in the finite time, too. When the described method is used it becomes an *instance of observation*. By application to 'itself' it becomes a *higher instance* of observation, now observing the trace of its previous instances. Thus the method of staging of the observed process will be used again. This method is applicable to all computing processes and as such it represents a contribution to the *dead lock* indication problem.

This sequence of ideas differs from the Cantor diagonal argument. Mainly it is achieved by the physical, especially by the thermodynamic and information (structure) type of our consideration respecting the *II. Principle of Thermodynamics* as the very principal root for methodological approaches of all types, both excluding and also enriching the mere empty logic.

2. Notion of Auto-Reference

Paradoxical claims (paradoxes, *noetical* paradoxes, *contradictions*, *antinomia*) have two parts - both parts are true, but the truth of one part denies the truth of the second part.

They can arise by not respecting the *metalanguage (semantic) level* - which is the higher level of our thinking about problems and the *language (syntactic) level* - which is the lower level of formulations of our 'higher' thoughts. Also they arise by not respecting a *double-level organization and description of measuring* - by not respecting the need of a *'step-aside' of the observer from the observed*. And also they arise by not respecting various *time clicks* in time sequences. As for the latter case they are in a contradiction with the *causality principle*. The common feature for all these cases is the *Auto-Reference* construction which itself, solved by itself, always states the requirement for ceasing the *II. Principle of Thermodynamics* and all its equivalents [8-14].

Let us us introduce the **Russel's criterion** for removing paradoxes[1]: **Within the flow of our thinking and speech we need and must distinguish between *two levels* of our thinking and expressing in order not to fall in a paradoxical claim *by mutual mixing and changing* them**.

These levels are the higher one, the metalanguage (semantic) level and the lower one, the language (syntactic) level. Being aware of the existence of these two levels we prevent ourselves from their mutual mixing and changing, we prevent ourselves from *application* our *metalanguage claims on themselves* but now on the language level or vice versa.

We must be aware that our claims about properties of considered objects are *created* on the higher level, rather richer both semantically and syntactically than the lower one on which we really express ourselves about these objects. The words and meanings of this lower (and 'narrower') level are common to both of them. Our speech is *formulated and performed* on the lower level describing here our 'higher' thoughts and on which the objects themselves have been

1 B. Russel, L. Whitehead, **Principia Mathematica**, 1910, 1912, 1913 and 1927.

described, defined yet too, of course from the higher level, but with the necessary (lower) limitations. (As such they are thought over on the higher level.) From this point of view we understand the various meanings (levels) of the same words. Then any mutual mixing and changing the metalanguage and language level or the autotereference (paradox, noetical paradox, contradiction, antinomium) is excluded.

2.1. Auto-Reference in Information Transfer, Self-Observation

In any *information transfer channel* $\mathcal{K}$ the *channel equation*

$$H(X) - H(X \mid Y) = H(Y) - H(Y \mid X) \tag{1}$$

it is valid [23]. This equation describes the mutual relations among *information entropies [(average) information amounts]* in the channel $\mathcal{K}$.

The quantities $H(X)$, $H(Y)$, $H(X \mid Y)$ and $H(Y \mid X)$ are the *input*, the *output*, the *loss* and the *noise* entropy.

The difference $H(X) - H(X \mid Y)$ or the difference $H(Y) - H(Y \mid X)$ defines the *transinformation* $T(X;Y)$ or the transinformation $T(Y;X)$ respectively,

$$H(X) - H(X \mid Y) \triangleq T(X;Y) \;=\; T(Y;X) \triangleq H(Y) - H(Y \mid X) \tag{2}$$

When the channel K transfers the information (entropy) $H(X)$, but now just at

the value of the entropy $H(X \mid Y)$, $H(X) = H(X \mid Y)$, then, necessarily, must be valid

$$T(X;Y) = 0 \quad \left[= H(Y) - H(Y \mid X) \right] \tag{3}$$

- For $H(Y \mid X) = 0$, we have $T(X;Y) = H(Y) = 0$.
- For $H(Y \mid X) \neq 0$ we have $H(Y) = H(Y \mid X) \neq 0$

In both these two cases the channel $\mathcal{K}$ operates as the *interrupted (with the absolute noise)* and the output $H(Y)$ is without any relation to the input $H(X)$ and, also, it doesn't relate to the structure of $\mathcal{K}$. This structure is expressed by the value of the quantity $H(X \mid Y)$. We assume, for simplicity, that $H(Y \mid X) = 0$.

From the (1)-(3) follows that the **channel** $\mathcal{K}$ **can't** transfer (within the same step p of its *transfer process*) such an information which describes its inner structure and, thus, it can't **transfer - observe** (copy, measure) **itself**. It it is valid both for the concrete information value and for the average information value, as well.

Any channel $\mathcal{K}$ can't transfer its own states considered as the input messages (within the same steps p). Such an attempt is the information analogy for the ***Auto-Reference*** *known from Logics and Computing Theory. Thus a certain 'step-aside' leading to a nonzero tranfer output, $H(Y)=H(X)-H(X\mid Y)>0$, is needed.*

2.2. Auto-Reference and Thermodynamic Stationarity

The transfer process running in an information transfer channel $\mathcal{K}$ is possible to be comprehended (modeled or, even, constructed) as the *direct* Carnot Cycle $\mathcal{O}$ [6, 8]. The relation $\mathcal{O}\cong\mathcal{K}$ is postulated. Further, we can imagine its observing method, equivalent to its 'mirror' $\mathcal{O}''\cong\mathcal{K}''$. This *mirror* $\mathcal{O}''$ is, at this case, the direct Carnot Cycle $\mathcal{O}$ as for its structure, but functioning in the *indirect, reverse* mode [6, 8].

Let us connect them together to a *combined heat cycle $\mathcal{O}\mathcal{O}''$ in such a way that the mirror (the reverse cycle $\mathcal{O}''$)* is gaining the message about the structure of the direct cycle $\mathcal{O}$. This message is (carrying) the information $H(X\mid Y)$ about the structure of the transformation (transfer) process ($\mathcal{O}\cong\mathcal{K}$) being 'observed'. *The mirror $\mathcal{O}''\cong\mathcal{K}''$ is gaining this information $H(X\mid Y)$ on its noise 'input' $H(Y''\mid X'')$ [while $H(X'')=H(Y)$ is its input entropy].*

The quantities ΔQ_W, ΔA and ΔQ_0 or the quantities $\Delta Q''_W$, $\Delta A''$ and $\Delta Q''_0$ respectively, define the information entropies of the information transfer realized (thermodynamically) by the *direct* Carnot Cycle $\mathcal{O}$ or by the *reverse* Carnot Cycle $\mathcal{O}''$ (the mirror) respectively (the *combined* cycle $\mathcal{O}\mathcal{O}''$ is created),

$$\begin{aligned} H(X)&=\frac{\Delta Q_W}{kT_W},\ \text{resp.}\ H(Y'')=\frac{\Delta Q''_W}{kT''_W}\\ H(Y)&=\frac{\Delta A}{kT_W},\ \text{resp.}\ H(X'')=\frac{\Delta A''}{kT''_W}\\ H(X\mid Y)&=\frac{\Delta Q_0}{kT_W},\ \text{resp.}\ H(Y''\mid X'')=\frac{\Delta Q''_0}{kT''_W}\end{aligned} \tag{4}$$

Our aim is to gain the *nonzero* output mechanical work ΔA^* of the combined heat cycle $\mathcal{O}\mathcal{O}''$, $\Delta A^*>0$. We want to gain nonzero information $H^*(Y^*)=\Delta A^*/kT_W>0$.

To achieve this aim, for the efficiencies η_{max} and η''_{max} of the both connected cycles $\mathcal{O}$ and $\mathcal{O}''$ (with the working temperatures $T_W=T''_W$ and $T_0=T''_0$, $T_W\geq T_0>0$), it must be valid that $\eta_{max}>\eta''_{max}$; we want the validity of the relation[2]

$$\Delta^* A=\Delta A-\Delta A''>0\ \left[\Delta A''=\Delta Q''_W-\Delta Q''_0\right] \tag{5}$$

2 We follow the proof of *physical and thus logical impossibility* of the construction and functionality of the Perpetuum Mobile of the *II.* and, equivalently [8], of the *I.* type.

When $\Delta Q_0 = \Delta Q''_0$ should be valid, then must be that $\Delta Q''_W < \Delta Q_W \Leftarrow (\eta_{max} > \eta''_{max})$] and thus it should be valid that

$$\begin{aligned} \Delta A^* &= \Delta Q_W \cdot \eta_{max} - \Delta Q''_W \cdot \eta''_{max} > 0 \ \text{ but} \\ \Delta Q_W \cdot \eta_{max} - \Delta Q''_W \cdot \eta''_{max} &= \Delta Q_0 - \Delta Q''_0 = 0 \end{aligned} \tag{6}$$

Thus the output work $\Delta A^* > 0$ *should be genarated* without any lost heat and by the *direct change* of the whole heat $\Delta Q_W - \Delta Q''_W$ but within the cycle $\mathcal{OO}''$. For $\eta_{max} < \eta_{max}$ the same heat $\Delta Q_W - \Delta Q''_W$ should be pumped from the cooler with the temperature T_0 to the heater with the temperature T_W *directly*, without any compensation by a mechanical work. We see that $\Delta A^* = 0$ is the reality.

Our combined machine $\mathcal{OO}''$ should be the *II. Perpetuum Mobile* in both two cases. Thus $\eta_{max} = \eta''_{max}$ must be valid (the heater with the temperature T_W and the cooler with the temperature T_0 are common) that

$$\eta_{max} = \eta''_{max} < 1 \ \text{ and then } \ \Delta Q_W = \Delta Q''_W \tag{7}$$

We must be aware that for $\eta_{max} = \eta''_{max} < 1$ the whole information entropy of the environment in which our (reversible) combined cycle $\mathcal{OO}''$ is running changes on one hand *by the value*

$$H(X) \cdot \eta_{max} = \frac{\Delta Q_W}{\mathrm{k}T_W} \cdot (1 - \beta) > 0, \ \beta = 1 - \eta_{max} = \frac{T_0}{T_W} \tag{8}$$

and on the other hand it is also changed *by the value* $-H(X) \cdot \eta_{max} = -\dfrac{\Delta Q_W}{\mathrm{k}T_W} \cdot (1-\beta)$

Thus it must be changed by the *zero* value

$$H^*(Y^*) = \frac{\Delta A^*}{\mathrm{k}T_W} = H(X) \cdot \eta_{max} - H(Y'') \cdot \eta''_{max} = H(X) \cdot (\eta_{max} - \eta_{max}) = 0 \tag{9}$$

The whole combined machine, or the thermodynamic system with the cycle $\mathcal{OO}''$ is, when the cycle $\mathcal{OO}''$ is seen, as a whole, in the *thermodynamic equilibrium*. (It can be seen as an unit, analogous to an interruptable operation in computing.)

Thus, the observation of the observed process $\mathcal{O}$ by the observing reverse process $\mathcal{O}''$ with the same structure (by itself), or the Self-Observation, is impossible in a physical sense, and, consequently, in a logical sense, too (see the Auto-Reference in computing).

Nevertheless, the construction of the Auto-Reference is describable and, as such, is recognizable, *decidable* just as a construction *sui generis*. It leads, necessarily, to the requirement of the *II.* Perpetuum Mobile functionality when the requirements (5) and (6) are sustained.

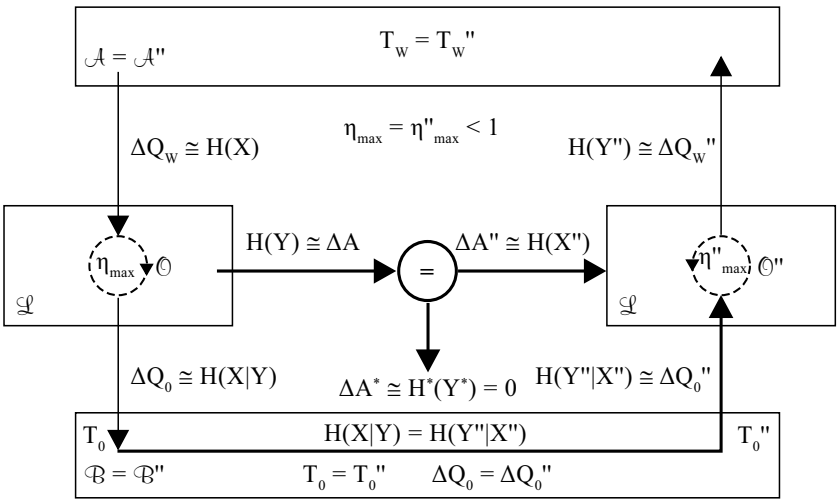

Figure 1. Stationarity of the double cycle $\mathcal{O}\mathcal{O}''$

(Note, that the Carnot Machine itself is, by its definition, a construction of the infinite cycle of the states of its working medium and as such is identifiable and recognizable.)

2.3. Gibbs Paradox and Auto-Reference in Observation and Information Transfer

Only just by a (thought) 'dividing' of an equilibrium system $\mathcal{A}$ by *diaphragms* [20], without any influence on its thermodynamic (macroscopic) properties, a non-zero difference of its entropy, before and after its 'dividing', is evidenced.

Let us consider a thermodynamic system $\mathcal{A}$ in volume V and with n matter units of ideal gas in the thermodynamic equilibrium. The *state equation* of $\mathcal{A}$ is $pV = nR\Theta$. For an elementary change of the *internal* energy U of $\mathcal{A}$ we have $\mathrm{d}U = nc_v\mathrm{d}\Theta$.

From the state equation of $\mathcal{A}$, and from the general *law of energy conservation* [for a (substitute) reversible exchange of heat δq between the system and its environment] we formulate the *I. Principle of Thermodynamics*, $\delta q = \mathrm{d}U + p\mathrm{d}V$

From this principle, and from *Clausius equation* $\Delta S \overset{\mathrm{Def}}{=} \frac{\Delta q}{\Theta}, \Delta q = c_v\Delta\Theta + \frac{R\Theta\Delta V}{V}, \Theta > 0$, follows that

$$S = n\int\left(c_v\frac{\mathrm{d}\Theta}{\Theta} + R\frac{\mathrm{d}V}{V}\right) = n\left(c_v\ln\Theta + R\ln V\right) + S_0(n) = \sigma\left(\Theta, V\right) + S_0(n) \tag{10}$$

Let us 'divide' the equilibrial system $\mathcal{A}$ in a volume V and at a temperature Θ, or, better said, the whole volume V (or, its whole state space) occupiable, and just occupied now by all its constituents (particles, matter units), with diaphragms (thin infinitely, or, 'thought' only), not affecting thermodynamic properties of $\mathcal{A}$ supposingly, to m parts $\mathcal{A}_i$, $i \in 1,m\}$, $m \geq 1$ with volumes V_i with matter units n_i. Evidently $n = \sum_{i=1}^{m} n_i$ and $V = \sum_{i=1}^{m} V_i$.

Let now $S_0(n)=0$ and $S_{0i}(n_i)=0$ for all i. For the entropies S_i of $\mathcal{A}_i$ considered individually, and for the change ΔS, when volumes V, V_i are expressed from the state equations, and for $p=p_i$, $\Theta=\Theta_i$ it will be gained that $\sigma_{[i]}=Rn_{[i]}\ln n_{[i]}$. Then, for $S_i=\sigma_i=n_i(c_v\ln\Theta+R\ln V_i)$ is valid, we have that

$$\begin{aligned}&\sum_{i=1}^{m}S_i=\sum_{i=1}^{m}\sigma_i=nc_v\ln\Theta+R\ln\left(\prod_{i=1}^{m}V_i^{n_i}\right),\\&\Delta S=S-\sum_{i=1}^{m}S_i=\sigma-\sum_{i=1}^{m}\sigma_i=\Delta\sigma=R\ln\frac{V^n}{\prod_{i=1}^{m}V_i^{n_i}}=-nR\sum_{i=1}^{m}\frac{n_i}{n}\ln\frac{n_i}{n}>0\end{aligned}\tag{11}$$

Let us denote the last sum as B further on, $B<0$. The quantity $-B$ expressed in (11) is information entropy of a source of messages with an alphabet $[n_1,n_2,n_m]$ and probability distribution $[n_i/n]_{i=1}^{m}$. Such a division of the system to m parts defines an information source with the information entropy with its maximum $\ln m$.

The result (11), $\Delta S=-nRB$, is a *paradox*, a contradiction with our presumption of not influencing a thermodynamic state of $\mathcal{A}$ by diaphragms, and, leads to that result that the heat entropy S (of a system in equilibrium) *is not* an extensive quantity. But, by the definition of the differential $\mathrm{d}S$, this *is not* true.

Due to this contradiction we must consider a non-zero integrating constants $S_0(n)$, $S_{0i}(n_i)$, in such a way, that the equation $\Delta S=(\sigma+S_0)-\sum_{i=1}^{m}(\sigma_i+S_{0i})=0$ is solvable for the system $\mathcal{A}$ and all its parts $\mathcal{A}_i$ by solutions $S_{0[i]}(n_{[i]})=-n_{[i]}R\ln\left(n_{[i]}/\gamma_{[i]}\right)$.

Then $S_{[i]}\triangleq S_{[i]}^{Claus}$, and we write and derive that

$$S^{Claus}=\sum_{i=1}^{m}S_i^{Claus}=\sum_{i=1}^{m}n_iR\ln\gamma_i=nR\ln\gamma\ \Rightarrow\ \gamma=\gamma_i;\ \Delta S=0.\tag{12}$$

Now let us observe an equilibrium, $S^*=S^{Claus}=S^{Boltz}=-kN\,B^*=-kN\ln N$.

Let, in compliance with the *solution* of Gibbs Paradox, the integration constant S_0 be the (change of) entropy ΔS which is added to the entropy σ *to figure out* the measured entropy S^{Claus} of the equilibrium state of the system A (the final state of Gay-Lussac experiment) at a temperature Θ. We have shown that without such correction, the less entropy σ is evidenced, $\sigma=S^{Claus}-\Delta S, \Delta S=S_0$.

Following the previuos definitions and results we have

$$\begin{aligned}&\Delta S=\frac{\Delta Q_0}{\Theta}=-nR\ln\frac{n}{\gamma},\\&\ln\gamma=\frac{\Delta S}{knN_A}+\ln n=\frac{\Delta S}{kN}+\ln N-\ln N_A,\ \gamma=N\ \Rightarrow\ \frac{\Delta S}{kN}=\ln N_A.\end{aligned}\tag{13}$$

By the entropy ΔS the 'lost' heat ΔQ_0 (at the temperature Θ) is defined.

Thus, our observation can be understood as an information transfer $\mathcal{T}$ in an information channel $\mathcal{K}$ with entropies $H(X)$, $H(Y)$, $H(X \mid Y)$ and $H(Y \mid X)$ in (4) but now bound physically; we have these information entropies per one particle of the observed system $\mathcal{A}$:

$$
\begin{aligned}
&\textit{input} \quad H(X) \overset{\text{Def}}{=} \frac{S*}{kN} = \ln \gamma = -B* = \ln N = -rB(r) \\
&\textit{output} \quad H(Y) \overset{\text{Def}}{=} \frac{\sigma}{kN} \triangleq -B^{Gibbs} = -B^{Boltz} = -B(r), \\
&\textit{loss} \quad H(X \mid Y) \overset{\text{Def}}{=} \frac{S_0}{kN}, \\
&\textit{noise} \; H(Y \mid X) \overset{\text{Def}}{=} 0 \;\; \text{for the simplicity;} \\
&H(X \mid Y) = -rB(r) - \left[-B(r)\right] = -B(r) \cdot (r-1) = \left(-B*\right) \cdot \frac{r-1}{r}, r \geq 1; \; \frac{1}{r} = \eta_{max}.
\end{aligned}
\tag{14}
$$

For a number m of cells of our railings in the volume V with $\mathcal{A}$, $m \leq N$ or for the accuracy r of this description of the 'inner structure' of $\mathcal{A}$ (a thought structure of V with $\mathcal{A}$) and for the number q of diaphragms creating our railings of cells and constructed in such a way that $q \in \langle 1, m-1 \rangle$, we have that $r = (N-1)/q$.

Our observation of the equilibrium system $\mathcal{A}$, including the *mathematical correction* for Gibbs Paradox, is then describable by the Shannon transfer scheme $[X, \mathcal{K}, Y]$ where

$$
H(X) = \frac{S^{Claus}}{kN}, \; H(X \mid Y) = \frac{S_0}{kN}, \; H(Y) = \frac{S^{Claus}}{kN}, \; H(Y \mid X) = \frac{\Delta S}{kN}. \tag{15}
$$

(However, a real observation process described in (15), equivalent to that one with $r = 1$, is impossible [6].)

We conclude by that, the diminishing of the measured entropy value about ΔS against $S*$ awaited, evidenced by **Gibbs Paradox,** ***does not originate in a watched system itself***. Understood this way, **it is a contradiction of a *gnozeologic character* based on not respecting *real* properties of any observation** [6]. And, this means the following.

The *minimal* accuracy of our description of the watched system $\mathcal{A}$ should be for $r = \infty$. In this case we should place $q = 0$ diaphragms, no railings is laid, $m = 1$, $q = 0$. 'We think nothing' about the 'inner structure' of the system $\mathcal{A}$, for in this case we are not outside of it (for the 'structure' measured - considered is 'created' by 0 diaphragms 'laid down' just from the outside). Thus we define an output information source Y, which is bound physically, and for which its (bound) information entropy is $H(Y) = -B^{Gibbs} = 0$. Then, the result of such 'observation' is 0, and the loss information entropy is

$$H(X \mid Y) = \frac{S^*}{kN} = \ln N = H(X).$$

If we consider $m > 1$ as the number of 'windows' being laid down over the measured (observed) system $\mathcal{A}$ *from the outside*, but now $m = 1$ and then $q = 0$, we resign the possibility of saying about the system $\mathcal{A}$ anything else than that it exists. The whole system $\mathcal{A}$ is now 'viewed' in the just one 'window' but now created by the volume V, the system $\mathcal{A}$ occupies, itself. This only one observation 'window' is 'pressed' to us by the system $\mathcal{A}$ itself and, by this way, 'we ourselves are identical' with it, 'we ourselves are' the system $\mathcal{A}$. We 'can move' within the volume V, but 'undivided' this time, and thus our measuring could be 'organized' now by the mediation of this only one window, but not laid down from the outside (by us). We are inside of this system (of the volume V) and our measuring is organized with the parameter $r = \infty$. We say that we are identical with the system for *we have now no 'step-aside' from it* $[H(Y) = H(X) - H(X \mid Y) = 0]$. This is the reason why we do not see it (from the outside) and, what we can think about it is nothing. In such a sense that we can't 'lay down' the diaphragms over the system ($q \neq 0$ is needed) and create on it our 'windows' ($m > 1$). Because we do not rule our measuring of the system $\mathcal{A}$ we do not 'divide' it by our, just from the outside laid, diaphragms (now $q = 0$) and, for this, we are not able to organize its measuring with the parameter $r < \infty$. 'We ourselves are' the system $\mathcal{A}$ and for this we can't, from the outside, see us, which means we can't see the system. (Or we can evidence its or our certain existence in this window.) The Auto-Reference is not possible; the measuring of the system $\mathcal{A}$ (from the outside **!**) is intended to be organized in the inside (**!**) of the system $\mathcal{A}$ itself.[3]

Our measuring also represents an observation of a certain phenomenon with the *degenerated* probability distribution $[n_i / n]_{i=1}^{1}$. The information amounts $i(X)$ and $H(X)$ of this phenomenon are equal to 0 and, due to this, our measuring organized by the measuring method with the parameter $m = 1$ or $r = \infty$ respectively will end with the result $H(Y) = 0$ $[H(Y) = \eta_{max} \cdot H(X H(X) = 0, \eta_{max} = 1/\infty]$.

For this all we need a certain 'step-aside' from the system $\mathcal{A}$ expressed by the value $r < \infty$, but, nevertheless with respecting properties of this 'step-aside', we do not fall into Gibbs Paradox [even when our railngs of diafragmas in the mode given $r \in \langle 1, \infty \rangle$ is laid down].

The ideal 'step-aside' is expressed by the value $r = 1$.

The *maximal* accuracy of our 'description', the accuracy of our watching of the system $\mathcal{A}$, is achieved for $r = 1$. Then $B(r) = B^*$ and for the output, the input and the loss information entropies it is valid that

$$H(Y) = H(X) = -B*, \; H(X \mid Y) = 0$$

Then we have the ideal 'step-aside' from the system $\mathcal{A}$. 'We have insight into its inside' and we can draw a layout of measuring its state without the distortion being given by the (at least

3 Our eye cannot see itself by itself only and, even more, it cannot see inside itself by itself only.

partial) transfer (energy) of its state into the organization of our measuring [in the form of the nonzero value of the information loss $H(X \mid Y)$ when $r \in (1,\infty\rangle$.]

This 'step-aside' enables us to set our measuring with the parameter $q = N-1, r = 1$. Now we know exactly what we measure, we know that we measure the status of the equilibrium thermodynamic system $\mathcal{A}$ and, by our 'step-aside' from it, we are able to check the precision (the organization) of the measuring method. Just this enables us to organize the measuring of the status of the system $\mathcal{A}$ without nonzero information loss $H(X \mid Y)$ in the other case being generated by the method itself (for $r \neq 1$), see the mentioned above gnozeological defect.

[But note that something quite different is the realization of the measuring method where the information losses are inevitable yet as the result of the validity of the *II*. Principle of Thermodynamics (and its equivalents [8]).]

3. Information Thermodynamic Concept Removing Auto-Reference

In the Auto-Reference case, *the whole combined machine $\mathcal{OO}''$ is a system in the equilibrium status. For this status we can introduce the term (quasi)**stationary status** in which the (infinitesimal) **part of heat is circulating**. Any round of this circulation is lasting the time interval Δt ; infinite, $\Delta t \to \infty$, for **not ideal model**, or, finite, $\Delta t \infty$, when the ideal model is used; then the part of heat cannot be the infinitesimal. With the exception of the II. Perpetuum Mobile functionality of this combined machine, which is not possible, see (5) and (6), only the opening of the system and an external activity, a **certain 'step-aside'** between the cycles $\mathcal{O}$ and $\mathcal{O}''$, moves it away (prevent it) from this status.*

Nevertheless, we sustain our wants of gaining the information (about) $H(X \mid Y)$ about the structure of the observed $\mathcal{O}$ (the transfer channel $\mathcal{K}$), we want the nonzero value ΔA^*, the nonzero information $H^*(Y^*) = \Delta A^* / \mathrm{k}T_W > 0$.

Then, necessarily, the mirror, the reverse Carnot Cycle $\mathcal{O}''$ (the transfer channel $\mathcal{K}''$) is to be constructed with that 'step-aside' (excluding that stationarity) from the observed $\mathcal{O} \cong \mathcal{K}$. Now we mean that the 'step-aside' of the *observing* process $\mathcal{O}$ from the *observed* process $\mathcal{O}''$ is realized by the difference $T_W - T''_W > 0$. Now, within this thermodynamic point of view, it is valid that $\Delta A'' < \Delta A''$ for $T_0 = T''_0$, $T''_W \triangleq T^*_0$

$$\Delta A'' = \Delta Q''_W \cdot \left(1 - \frac{T_0}{T''_W}\right) = \Delta Q_W \cdot \frac{T''_W}{T_W} \cdot \left(1 - \frac{T_0}{T''_W}\right) \tag{16}$$

Then, for the whole information amount $\Delta A^* / \mathrm{k}T_W$ of our combined cycle it is valid that

$$\frac{\Delta A^*}{\mathrm{k}T_W} = H(Y'') - H(Y'') \cdot \beta^* = H(Y'') \cdot \left(1 - \frac{T_0^*}{T_W}\right) = H(Y'') \cdot \left[1 - \frac{H(X \mid Y)}{H(X'' \mid Y'')}\right] \tag{17}$$

The structure $H(X \mid Y)$ of the observed transfer (channel, process) $\mathcal{O} \cong \mathcal{K}$ is measurable with the 'step-aside' only, created now by different temperatures ($T_W > T''_W$). The result is

$$\frac{\Delta A^*}{\mathrm{k}T_W} > 0, \left[\frac{\Delta A^*}{\mathrm{k}T_W} \cong f\left[H(X \mid Y)\right] > 0\right] \tag{18}$$

Following (5), (6) and (9) the Auto-Reference arises just when

$$T_W = T''_W \ \left[\Rightarrow H(Y) = 0\right] \tag{19}$$

Now we will describe, in the *information thermodynamic way*, the problem of a Self-Observation or the Auto-Reference problem and will draw a method of its ceasing.

For we want to have the information $H(X \mid Y)$ describing the structure of the transfer process $O \cong T$ which we observe we need gain a nonzero value (difference) ΔA^* and, consequently, a nonzero information $H^*(Y^*)$,

$$H^*(Y^*) = \frac{\Delta A^*}{\mathrm{k}T_W} > 0 \tag{20}$$

Then we need a 'mirror', the reverse Carnot Cycle $\mathcal{O}'' \cong T''$ (or the relevant transfer channel $\mathcal{K}''$) would be constructed in such a way that the mentioned 'step-aside' from the observed transfer channel $\mathcal{K}$ was respected. [It is the 'step-aside' of the observing process ($\mathcal{O}''$, $\mathcal{T}''$) from the observed process ($\mathcal{O}$, $\mathcal{T}$); also we can consider a computing process $\vec{\kappa}$ and its description - the program $\vec{\eta}$, and its observation, see later].

Now the required 'step-aside' is realized by the temperature difference $T_W - T''_W > 0$. Thus now we consider, within the frame of our thermodynamic approach, that $\Delta A'' < \Delta A''$ for $T_0 = T''_0$ and then, under the condition $\Delta Q_0 = \Delta Q''_0$, it will be valid for the cycles $\mathcal{O}$, $\mathcal{O}''$ and $\mathcal{O}\mathcal{O}''$ that

$$\begin{aligned} \Delta Q''_W &= \Delta A'' + \Delta Q''_0 \\ &= \Delta Q''_W \cdot \eta''_{max} + \Delta Q_W \cdot \left(1 - \eta_{max}\right) \\ &= \Delta Q''_W \cdot \eta''_{max} + \Delta Q_W - \Delta Q_W \cdot \eta_{max} \end{aligned} \tag{21}$$

but also it is valid that

$$\begin{aligned} \Delta Q_W &= \Delta Q''_W \cdot \eta''_{max} + \Delta Q''_W \cdot \left(1 - \eta''_{max}\right) \\ &= \Delta Q''_W \cdot \eta''_{max} + \Delta Q''_W - \Delta Q''_W \cdot \eta''_{max} \end{aligned} \tag{22}$$

From the proposition $\Delta Q_0 = \Delta Q''_0$ [from relations (21) and (22)] pro $\Delta Q''_W$ follows that

$$\Delta Q_W \cdot \left(1-\eta_{max}\right) = \Delta Q''_W \cdot \left(1-\eta''_{max}\right) \tag{23}$$

With the denotation $\beta \triangleq (1-\eta_{max}, \beta'' \triangleq (1-\eta''_{max})$ we write

$$\begin{aligned} &\Delta Q_W \cdot \beta = \Delta Q''_W \cdot \beta'' \\ &\frac{\Delta Q_W}{\Delta Q''_W} = \frac{T_W}{T''_W} = \frac{\frac{T_0}{T''_W}}{\frac{T_0}{T_W}} = \frac{\beta''}{\beta} \\ &\qquad \Rightarrow \\ &\Delta Q''_W = \Delta Q_W \cdot \frac{\beta}{\beta''} = \Delta Q_W \cdot \frac{T''_W}{T_W},\ \Delta Q''_W < \Delta Q_W \end{aligned} \tag{24}$$

Within the cycles $\mathcal{O}$ and $\mathcal{O}''$ the following relations are valid,

$$\begin{aligned} &\frac{T_W - T_0}{T_W} > \frac{T''_W - T_0}{T''_W},\ T_W > T''_W \\ &\frac{T_0}{T''_W} > \frac{T_0}{T_W} \ \text{ a tedy } \ \frac{Q_0}{Q''_W} > \frac{Q_0}{Q_W},\ Q_W > Q''_W \\ &\qquad \Rightarrow \\ &\frac{H\left(Y'' \mid X''\right)}{H\left(Y''\right)} > \frac{H\left(X \mid Y\right)}{H\left(X\right)} \end{aligned} \tag{25}$$

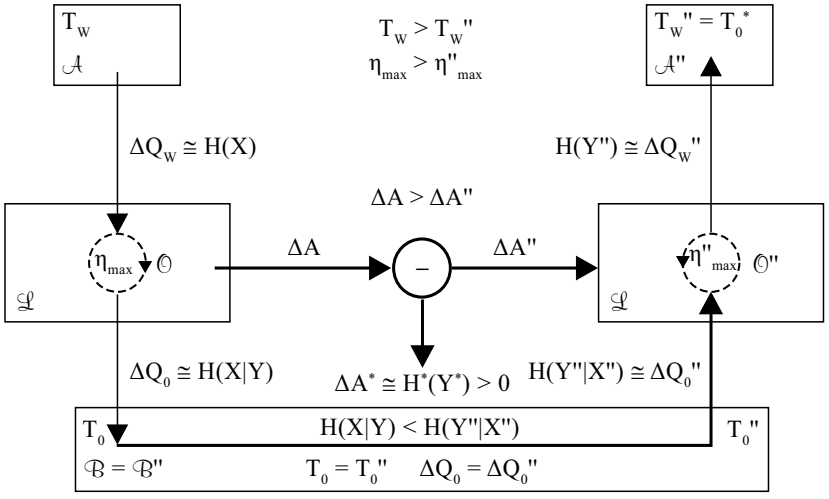

Figure 2. The concept for ceasing the Auto-Reference

By (23) and (24) for $\Delta A''$ it is valid in the cycle $\mathcal{O}''$ that

$$\begin{aligned}\Delta A'' &= \Delta Q''_W \cdot \left(1-\frac{T_0}{T''_W}\right) = \Delta Q_W \cdot \frac{T''_W}{T_W} \cdot \left(1-\frac{T_0}{T''_W}\right) = \\ &= \Delta Q_W \cdot \left(\frac{T''_W}{T_W} - \frac{T_0}{T_W}\right) = \mathrm{k} \cdot H(X) \cdot \left(T''_W - T_0\right) \\ &= \mathrm{k} \cdot H(X) \cdot T''_W \left(1-\frac{T_0}{T'_W}\right) = \mathrm{k} \cdot H(X) \cdot T''_W (1-\beta'') = \mathrm{k} \cdot T''_W \cdot H(Y'')\end{aligned} \tag{26}$$

and, further, for ΔA in the cycle $\mathcal{O}$ we have

$$\Delta A = \mathrm{k} \cdot H(X) \cdot T_W (1-\beta) = \mathrm{k} \cdot H(X) \cdot T_W \left(1-\frac{T_0}{T_W}\right) \tag{27}$$

and thus, for the cycles $\mathcal{O}''$ and $\mathcal{O}$ it is valid that

$$\frac{\Delta A''}{\mathrm{k}T''_W} = H(X) \cdot \left(1-\frac{T_0}{T''_W}\right) = H(X) \cdot (1-\beta'') = H(X) \cdot \eta''_{max} \tag{28}$$

$$\frac{\Delta A}{\mathrm{k}T_W} = H(X) \cdot \left(1-\frac{T_0}{T_W}\right) = H(X) \cdot (1-\beta) = H(X) \cdot \eta_{max}$$

For the whole work ΔA^* of the combined cycle $\mathcal{O}\mathcal{O}''$ we have

$$\Delta A^* = \Delta A - \Delta A'' = \left[\mathrm{k}T_W \cdot H(X) \cdot (1-\beta) - kT''_W \cdot H(X) \cdot (1-\beta'')\right] > 0 \tag{29}$$

Then, for the *whole change* of the thermodynamic entropy within the combined cycle $\mathcal{O}\mathcal{O}''$ (measured in information units *Hartley*, *nat*, *bit*) and thus for the change of the whole information entropy $H^*(Y^*)$ $[\triangleq H_C]$, following the relation (29), it is valid

$$\begin{aligned}H^*(Y^*) = \frac{\Delta A^*}{\mathrm{k}T_W} &= H(X) \cdot \left[(1-\beta) - \frac{T''_W}{T_W} \cdot (1-\beta')\right] \\ &= H(X) \cdot \left(1 - \frac{T_0}{T_W} - \frac{T''_W}{T_W} + \frac{T_0}{T_W}\right) = H(X) \cdot \left(1-\frac{T''_W}{T_W}\right)\end{aligned} \tag{30}$$

It is valid, for ΔA^* is a *residuum work* after the work ΔA has been performed at the temperature T_W. Evidently, the sense of the symbol T''_W (within the double cycle $\mathcal{OO}''$ and when $\Delta Q_0 = \Delta Q''_0$) is expressible by the symbol T_0^*, which is possible, for the working temperatures of the whole cycle $\mathcal{OO}''$ are T_W and $T''_W = T_0^*$. The relation (30) expresses that fact that the double cycle $\mathcal{OO}''$ is the direct Carnot Cycle just with its working temperatures $T_W > T''_W = T_0^*$. In the double cycle $\mathcal{OO}''$ it is valid that

$$\begin{aligned}
&\beta'' = \frac{\Delta Q''_0}{\Delta Q''_W} = \frac{\frac{\Delta Q''_0}{T''_W}}{\frac{\Delta Q''_W}{T''_W}} = \frac{H(Y''\,|\,X'')}{H(Y'')} = \frac{T_0}{T''_W},\ T''_W = T_0^*,\ \text{cyklus}\ \mathcal{O}'' \\
&\beta = \frac{\Delta Q_0}{\Delta Q_W} = \frac{\frac{\Delta Q_0}{T_W}}{\frac{\Delta Q_W}{T_W}} = \frac{H(X\,|\,Y)}{H(X)} = \frac{T_0}{T_W},\ \text{cyklus}\ \mathcal{O} \\
&\frac{\beta}{\beta''} = \frac{T''_W}{T_W} = \frac{T_0^*}{T_W} \triangleq \beta^*
\end{aligned} \tag{31}$$

and then, by (30) a (31) is writable that

$$\frac{\Delta A^*}{\mathrm{k}T_W} = H(X)\cdot\left(1-\beta^*\right) = H(X)\cdot\left[1-\frac{H(X\,|\,Y)\cdot H(Y'')}{H(Y''\,|\,X'')\cdot H(X)}\right] > 0 \tag{32}$$

It is ensured by the propositions $T_W > T''_W$, $T''_0 = T_0$ and also by that fact, that the loss entropy $H(X\,|\,Y)$ is described and given by the heat $\Delta Q_0 = \Delta Q''_0$. But by our combined cycle $\mathcal{OO}''$ it is valid too that

$$H(X) = \frac{\Delta Q_W}{\mathrm{k}T_W} = \frac{\Delta Q''_W}{\mathrm{k}T''_W} = H(Y'')\ \left[= \frac{\Delta Q''_W}{\mathrm{k}T_0^*}\right] \tag{33}$$

and by both parts of (4) we have

$$\frac{H(X\,|\,Y)}{H(Y''\,|\,X'')} = \beta^* < 1 \tag{34}$$

For the whole information entropy $\Delta A^*/\mathrm{k}T_W$ (the whole thermodynamic entropy $\mathcal{S}_C$ in information units) and by following the previous relations also it is valid that

$$\begin{aligned}\frac{\Delta A^{*}}{\mathrm{k}T_{W}} &= H(Y'') - H(Y'')\cdot\beta^{*} = H(Y'')\cdot\left(1-\frac{T_{0}^{*}}{T_{W}}\right)\\ &= H(Y'')\cdot\left[1-\frac{H(X\mid Y)}{H(X''\mid Y'')}\right]\end{aligned} \tag{35}$$

And thus, the structure of the information transfer channel $\mathcal{K}$ (expressed by the quantity $H(X\mid Y)$) is measurable by the value $H^{*}(Y^{*})$ from (20), (32) and (35). Symbolicaly, we can write, using a growing function f,

$$H^{*}(Y^{*}) = \frac{\Delta A^{*}}{\mathrm{k}T_{W}} \cong f\left[H(X\mid Y)\right] > 0 \tag{36}$$

The cycles $\mathcal{O}$, $\mathcal{O}''$ and $\mathcal{O}\mathcal{O}''$ are the Carnot Cycles and thus, from their definition and construction they are, imaginatively[4] in principle, the **infinite** cycles; in each of them the following *criterion of an infinite cycle* (see [14]) it is valid inevitably,

$$T\left(X^{[\cdot]};Y^{[\cdot]}\right) = H\left(X^{[\cdot]}\right) - H\left(X^{[\cdot]}\mid Y^{[\cdot]}\right) = H\left(Y^{[\cdot]}\right) > 0 \ \text{ and } \ \Delta S_{\mathcal{L}}^{[\cdot]} = 0 \tag{37}$$

The construction of the cycle $\mathcal{O}\mathcal{O}''$ enables us to recognize that the *infinite cycle* $\mathcal{O}$ is running; the unsolvable case of the Auto-Reference in $\mathcal{O}\mathcal{O}''$ occurs just when and only when

$$T_{W} = T_{W}'', \left[\left[T_{W} = T_{W}''\right] \Rightarrow \left[H^{*}(Y^{*}) = 0\right]\right] \tag{38}$$

(Thus, when we contemporarily await anything else than the zero output or an output which is not relevant to the given input, by this only awaiting, we require a construction of the Perpetuum Mobile functionality.)

Our observation of the process $\mathcal{O}$ now being considered as a model or as the realization of the computing process $\vec{\kappa}$ in a certain Turing Machine TM or an information transfer $\mathcal{T}$ in a certain channel $\mathcal{K}$ - the measuring of a transferring structure expressed by the value of the quantity $H(X\mid Y)$, is thus possible but only by another process with a certain 'step-aside' from the observed process $\mathcal{O}$ in-built, with a certain 'mirror' $\mathcal{O}''\cong\mathcal{T}''$, with the aids of another transferring structure expressed by the value of the quantity $H(X''\mid Y'')$.

From the both processes, the cycles $\mathcal{O}$ and $\mathcal{O}''$, the whole combined and one cycle $\mathcal{O}\mathcal{O}''$ is created to be modeling the whole and one transfer channel $\mathcal{K}\mathcal{K}''$ in which the observed, the measured property, the value $H(X\mid Y)$ in the given 'time click' p (the click, the interval) expressing the

4 When an infinite reserve of energy would exist.

structure of the channel $\mathcal{K}$ is one of the ingoing (input) messages (informations). The resulting double cycle $\mathcal{OO}''$ performs as a direct Carnot Cycle with the working temperatures T_W and T^*_0, $T_W > T^*_0$, $T''_W = T^*_0$.

The required 'step-aside' is created by the difference $T_W - T^*_0 > 0$. The whole information entropy (the thermodynamic entropy in information units) of the whole isolated system in which our double cycle $\mathcal{OO}''$ is running, the whole output information $H^*(Y^*)$ of this double cycle $\mathcal{OO}''$ is then a certain function $f(\cdot)$ of the *measure* $H(X \mid Y)$ of the structure being measured (observed) and as such, of the value of the argument $H(X \mid Y)$ from the relation (36).

Remark: Also the following consideration is possible: We use the change $H(Y)$ of the output entropy of the observed process $\mathcal{O}$ as its reaction to a change $H(X)$ of the input entropy and just by the evidenced $H(X \mid Y)$. Our measuring is then characterized, in the same sense as in (36) and (37), by the whole value

$$H(X)\cdot\left(1-\beta\right)-H(X)\cdot\left(1-\beta''\right)=H(X)\cdot\frac{T_0}{T''_W}\cdot\left(1-\frac{T''_W}{T_W}\right)>0$$

Now it is obvious that the transferring (copying, measuring, observing) of the structure of the channel $\mathcal{K}$ is possible then and only then a certain structural 'step-aside' between the observed object (the transferred or the measured structure, now and here the structure of a transfer channel $\mathcal{K}$) and the observing process (the channel $\mathcal{K}''$ with the transferring process $\mathcal{T}''$) expressed by the nonzero and positive values of the difference (2) is possible.

By the term 'step-aside' of the **observing** computing process (let us denote it as $\vec{\kappa}''$) from the **observed** computing process (let us denote it as $\vec{\kappa}$) we understand a **time delay** between them, better said, it is a **tracing**[5] of the observed computing process $\vec{\kappa}$.

4. Turing Computing, Auto-Reference and Halting Problem

Turing Machine (TM) is driven by a *program* which is interpreted by its *Control Unit* (CU_{TM}). The Control Unit CU_{TM} is a *finite automaton* (Mealy's or Moore's *sequential machine*). *The program* for the TM consists of the finite sequence $\vec{\eta}$ of instructions $\eta_{[\cdot]}$,

$$\vec{\eta}=\left(\eta_q\right)_{q=1}^{q\in\mathbb{N}}=\left[\left(s_i,x_k,s_j,y_l,D\right)_q\right]_{q=1}^{q\in\mathbb{N}},\ \left|\vec{\eta}\right|\in\mathrm{N} \tag{39}$$

Each of these instructions describes an overwriting rule of a *regular grammar* [15, 19, 21],

5 In the programmers' manner or language: listing, cross-reference, memory dump.

$$s_i \to (x_k, y_l, D) s_j \tag{40}$$

performed in the given step (time, moment) p, $p \in \mathbb{N}$, of the TM 's activity;

- s_i is the i -th *nonterminal symbol* of the regular grammar, or, respectively, it is a *status* of the CU_{TM} within the actual step $p \in \mathbb{N}$ of the TM 's activity,
- x_k is an *input terminal symbol* being read from the *input-output tape* of the TM within the actual step p of the TM 's activity,
- y_l is an *output terminal symbol* by which the CU_{TM} overwrites the symbol x_k which has been read (in the actual step p of the TM 's activity),
- s_j is the *successive* status of the CU_{TM}, given by the instruction for the following step $p+1$ of the CU_{TM}.

Within the actual step p of the TM 's activity the CU_{TM} is changing its status to s_j [this change is based on the status s_i, and the symbol x_k has been read $(s_i \overset{x_k}{\to} s_j)$], and is performing the transformation

$$x_k \to y_l \tag{41}$$

on the scanned (actual) *position* of the input-output tape,

- D determines the *moving direction* of the *read-write head* of the CU_{TM} after the symbol y_l has been recorded [in the status s_{j_p} (s_{j_p} denotes s_j for the step p) used further on as the following one, $s_{j_p} \overset{\text{Def}}{=} s_{i_{p+1}}$], $D \in \{L, R\}$.

The value L or R of the symbol D determines the *left slip* or the *right slip* from the actual position on the input-output tape to its (left or right) neighbor after the transformation x_k to y_l has been performed.

The *oriented edge* of the *transition graph* of the CU_{TM} (the finite automaton) is described by the symbol $s_i \overset{(x_k, y_l, D)}{\to} s_j$. The TM 's activity generates a sequence of the instructions *having been performed* in steps p, $\left[(s_{i_p}, x_{k_p}, s_{j_p}, y_{l_p}, D_p)\right]_{p=1}^{p=p_{\text{last}}}$,

$$s_{i_p} \xrightarrow{(x_{k_p}, y_{l_p}, D_p)} s_{j_p}, \text{ further on } \quad s_{j_p} = s_{i_{p+1}} \tag{42}$$

(the edge of the oriented transition graph of the CU_{TM} in the step p), by which the *computing process* $(\vec{\kappa})$ has gone through (from the first step $p=1$ till, *for this while*, the last step $p = p_{\text{last}}$ of the

TM 's activity). They are also the *overwriting rules* of the regular grammar, being performed within each step $p, p \geq 1$,

$$s_{i_p} \rightarrow \left(x_{k_p}, y_{l_p}, D_p \right) s_{j_p}, \; s_{j_p} = s_{i_{p+1}} \tag{43}$$

By this way a *regular language* of the *words* (x_{k_p}, y_{l_p}, D_p) or, respectively, a regular language of the instructions $(s_{i_p}, x_{k_p}, s_{j_p}, y_{l_p}, D_p)$ having been performed is defined. This second regular language is describable by the rules (of a regular grammar)

$$S_{i_p} \rightarrow \left(s_{i_p}, x_{k_p}, s_{j_p}, y_{l_p}, D_p \right) S_{j_p}, \; S_{j_p} \overset{\text{Def}}{=} S_{i_{p+1}} \tag{44}$$

being applied in each step $p \geq 1$ of the *TM* 's activity. Thus, this language is to be acceptable by a certain finite automaton with n states $S_{[\cdot]}$.

When this language is ***infinite***[6] *(the infinite chain of instructions of the finite length), such its word*

$$\left[\left(s_{i_1}, x_{k_1}, s_{j_1}, y_{l_1}, D_1 \right), ..., \left(s_{i_p}, x_{k_p}, s_{j_p}, y_{l_p}, D_p \right) \right]_{p=l} \tag{45}$$

of the length l exists that for that finite automaton [with n states $S_{[\cdot]}$ *and the transition rules (42) or (44)] the* ***Pumping Lemma*** *[19, 21]* ***it is valid***

$$n \leq l < 2n \tag{46}$$

4.1. Auto-Reference in Turing Computing

Although any instruction of the Turing Machine *TM* describes one step of the *computing process* in this *TM*, it is considerable as a *description (of one step) of an information transfer process* running in a certain transfer channel $\mathcal{K}$; we postulate the relation $TM \cong K$. The computing process in the *TM* is, also, a transfer process in a channel $\mathcal{K}$. For $\mathcal{K} \cong \mathcal{O}$ it is valid that $TM \cong \mathcal{O}$.

In each step $p > 1$ of its activity, the $TM \cong \mathcal{K}$ is accepting *its own configuration from the previous step* $p-1$ *as its input*, includes its contemporary status ($s_{i_p} = s_{j_{p-1}}$) and generates its status $[s_{i_{(p+1)}}]$ and the configuration for the next step $p+1$, etc.[7]

6 Better said, having the *arbitrary* (but finite) *length*.

7 $(s_{i_p}, x_{k_p}, s_{j_p}, y_{l_p}, D_p)(\vec{\sigma}_p, s_p, \vec{\rho}_p) \rightarrow (\overrightarrow{\sigma_{p+1}}, s_{i_{p+1}}, \overrightarrow{\rho_{p+1}})$, see further.

Similarly it is valid for the configurations (denoted now by X_p and Y_p), see further on.

For each $p \geq 1$ we consider the actual instances of the stochastic quantities[8] X, Y,

$$\begin{aligned} &X \triangleq X_p, Y \triangleq Y_p;\ X \mid Y \triangleq X_p \mid Y_p, Y \mid X \triangleq Y_p \mid X_p;\ Y_p = X_{p+1} \\ &X_p \mid X_{p+1} \triangleq X_p \sqsubseteq X_p X_{p+1},\ X_{p+1} \mid X_p \overset{\Delta}{=} X_{p+1} \sqsubseteq \left(X_p X_{p+1}\right)^{-1} \end{aligned} \tag{47}$$

In any step p of the TM 's activity its own configurations $(\vec{\sigma}_p, s_p, \vec{\rho}_p)$ - *members of the sequence* - of the **computing process** $\vec{\kappa} \overset{\text{Def}}{=} [(\vec{\sigma}_p, s_p, \vec{\rho}_p)]_{p=1}^{p\ldots}$, **can be considered as follows**;

- let now the stochastic quantity X_p be realized by the chain

$$\left(\vec{\sigma}_p, s_p, \vec{\rho}_p\right) \in \mathbf{T}^* \times \mathbf{S} \times \mathbf{T}^*;\ p = 1,\ \vec{\sigma}_1 = \vec{\varepsilon}\ \ a\ \ \vec{\rho}_1 = \vec{\xi},\ s_p = s_0 \tag{48}$$

- let now the stochastic quantity Y_p be realized by the chain

$$\left[\vec{\sigma}_{p+1}, s_{p+)}, \vec{\rho}_{p+1}\right] \in \mathbf{T}^* \times \mathbf{S} \times \mathbf{T}^* \tag{49}$$

Then, the computing process in the $TM \cong \mathcal{K}$ is describable informationally,[9,10]

$$\begin{aligned} H(X) &\triangleq H(X_p) = H\left(\overrightarrow{\sigma_p}, s_p, \overrightarrow{\rho_p}\right) \\ H(Y) &\triangleq H(Y_p) = H\left[X_{p+1}\right] = H\left[\overrightarrow{\vec{\sigma}_{p+1}}, s_{p+1}, \overrightarrow{\rho_{p+1}}\right] \\ H(X \mid Y) &\triangleq H\left(X_p \mid Y_p\right) = H\left[X_p \mid X_{p+1}\right],\ H\left(Y \mid X\right) \triangleq H\left(Y_p \mid X_p\right) = H\left[X_{p+1} \mid X_p\right] \\ T(X;Y) &\triangleq T\left(X_p; Y_p\right) = H\left(X_p\right) - H\left[X_p \mid X_{p+1}\right] \\ &= H\left(\overrightarrow{\sigma_p}, s_p, \overrightarrow{\rho_p}\right) - H\left[\left(\overrightarrow{\sigma_p}, s_p, \overrightarrow{\rho_p}\right) \mid \left[\overrightarrow{\sigma_{p+1}}, s_{p+1}, \overrightarrow{\rho_{p+1}}\right]\right] \\ T(Y;X) &\triangleq T\left(Y_p; X_p\right) = H\left(X_{p+1}\right) - H\left[X_{p+1} \mid X_p\right] \\ &= H\left[\overrightarrow{\sigma_{p+1}}, s_{p+1}, \overrightarrow{\rho_{p+1}}\right] - H\left[\overrightarrow{\sigma_{p+1}}, s_{p+1}, \overrightarrow{\rho_{p+1}}\right] \mid \left(\overrightarrow{\sigma_p}, s_p, \overrightarrow{\rho_p}\right)\Big] \end{aligned} \tag{50}$$

The Auto-Reference arises with the following description of the computing (transfer, observation) process when, e.g., for a certain $p \geq p^* \geq 1$,

---

8 '$\sqsubseteq$' now denotes the *substring* from the beginning of the string.

9 The transitions are given by the η_{q_p} called by $X_p X_p \mid X_{p+1} \rightarrow \eta_{q_p}, [\eta_{q_p}^{-1}(X_{p+1}) = X_p]$.

10 $\eta_{q_p}^{-1}(X_{p+1}) = X_p$ - comparison of the structures of the X_p and the X_{p+1} (in $X_p \mid X_{p+1}$).

$$
\begin{aligned}
&H(X_p) - H\left[X_p \mid X_{p+1}\right] = H(X_{p+1}) - H\left[X_{p+1} \mid X_p\right],\ p \geq 1 \\
&H(X_p) = H\left(X_p \mid X_{p+1}\right);\ X_p = X_p \sqsubseteq X_{p+1},\ X_{p+1} = \varepsilon \\
&H(X_{p+1}) = H\left(X_{p+1} \mid X_p\right),\ \left[H\left[X_{p+1} \mid X_p\right] = 0\right]
\end{aligned}
\tag{51}
$$

This way of considerations 'constructs' the TM 's infinite cycle from the programmer's point of view, Self-Observation in an information point of view and a stationary status from the thermodynamics point of view.

In any case a 'step-aside' to gain something else than the zero output is required.

By the 'step-aside' of the *observing* computing process from the *observed* computing process we mean a *time delay* between those two processes or, better said, a *staging of the trace* of the observed process.

4.2. Halting Problem as Auto-Reference

Now we are considering a certain TM (the observed machine) being driven by a program $\vec{\eta}$ and working with a certain input word $\vec{\xi}$. Let this activity be described by the word $[\mathrm{d}(TM)]$.

Let us consider that the TM with the input word $\vec{\xi}$

- **halts**, $\boldsymbol{HALT}_{\mathbf{TM}}$, whether the word $\vec{\xi}$ *is accepted* or *rejected*,

$$
\boldsymbol{HALT}_{\mathbf{TM}} = \left\{\boldsymbol{HALT}_{\mathbf{TM}}^{\mathrm{Accept}} \cup \boldsymbol{HALT}_{\mathbf{TM}}^{\mathrm{Reject}}\right\} \tag{52}
$$

- **does not halt**, $\boldsymbol{LOOP}_{\mathbf{TM}}^{\infty}$ (the TM 's infinite cycle)

Let us construct the three new Turing Machines M_1, M_2 and M_3 as follows [18][11]

- M_1 works with the input word $[\overrightarrow{\mathrm{d}(TM)} * \vec{\xi}]$ in that way that
- **halts**, $\boldsymbol{HALT}_{\mathbf{M_1}}^{\mathrm{Accept}}$,
- **stops**, $\boldsymbol{HALT}_{M_1}^{\mathrm{Reject}}$,

$$
\boldsymbol{HALT}_{M_1}^{\mathrm{Accept}} \Leftarrow \boldsymbol{HALT}_{TM},\qquad \boldsymbol{HALT}_{M_1}^{\mathrm{Reject}} \Leftarrow \boldsymbol{LOOP}_{TM}^{\infty} \tag{53}
$$

- M_2 **modifies the activity of the** M_1 in that way, that the input word which is being worked with is $[\overrightarrow{\mathrm{d}(TM)} * \vec{\xi}]$ and
- **halts**, $\boldsymbol{HALT}_{M_2}$,

11 Minski's proof for the undecidability of the Halting Problem (Entscheidungsproblem type).

- **does not halt**, $LOOP^{\infty}_{M_2}$,

$$\begin{aligned} HALT_{M_2} &\Leftarrow LOOP^{\infty}_{TM} \left[\Rightarrow HALT^{\text{Reject}}_{M_1}\right] \\ LOOP^{\infty}_{M_2} &\Leftarrow HALT_{TM} \left[\Rightarrow HALT^{\text{Accept}}_{M_1}\right] \end{aligned} \tag{54}$$

- M_3 **is an 'extension' of the** M_2 : it doubles its own input word $[\overrightarrow{\mathrm{d}(TM)}]$ into $[\overrightarrow{\mathrm{d}(TM)} * \overrightarrow{\mathrm{d}(TM)}]$ and gives it to the input (of its sub-machine) M_2 and

- **halts**, $HALT_{M_3}$,

- **does not halt**, $LOOP^{\infty}_{M_3}$,

$$\begin{aligned} HALT_{M_3} \equiv HALT_{M_2} &\Leftarrow LOOP^{\infty}_{TM} \left[\Rightarrow HALT^{\text{Reject}}_{M_1}\right] \\ LOOP^{\infty}_{M_3} \equiv LOOP^{\infty}_{M_2} &\Leftarrow HALT_{TM} \left[\Rightarrow HALT^{\text{Accept}}_{M_1}\right] \end{aligned} \tag{55}$$

- But, when the machine $M_3 \equiv TM$ accepts the description $\overrightarrow{d(M_3)}$, thus it is valid that $\overrightarrow{d(M_3)} \equiv \overrightarrow{\mathrm{d}(TM)}$, then

$$\begin{gathered} \left[HALT_{TM} \Leftarrow LOOP^{\infty}_{TM}\right] \wedge \left[LOOP^{\infty}_{TM} \Leftarrow HALT_{TM}\right] \\ \equiv \\ HALT_{TM} \Leftrightarrow LOOP^{\infty}_{TM} \end{gathered} \tag{56}$$

This result (56) is the **contradiction**. It is the consequence of the Cantor diagonal argument having been used carrying the Auto-Reference to the sequence of the machines (TM, M_1, M_2, M_3), or, respectively, to the sequence of the machines (TM, M_3),

$$\left(\underline{TM}, M_3 \equiv \underline{TM}\right) \tag{57}$$

and is **leading us to that opinion that the proposition about the decidability of the Halting Problem** (recognizing the $LOOP^{\infty}_{TM}$ state) **is not right**.

In any given step $p \geq 1$ the machine TM is deciding about itself (it is working with the description of its own actual status), it is the channel 'transferring' its own structure, it is the Self-Observer. Thus it is in a stationary status in the thermodynamic point of view.

Within this point of view, we can envisage two identical, but reversed mutually, ideal Carnot Cycles connected together. In this sense, these two machine $\mathcal{O}$ *and* $\mathcal{O}''$ *create the equilibrium system* OO''*, in which we introduce the term* ***stationary status****.*

Within such a system the (infinitesimal) ***part of heat is circulating*** *through the whole combined machine* $\mathcal{OO}''$*. Let this fact be now the* ***thermodynamic model of the infinite cycle*** *being started by the Self-Observation, by the Auto-Reference action (56), (57); one run is like an uninterruptable operation.*

The *recursive call* of the function $\overrightarrow{\mathrm{d}(TM)}$ (of the machine TM) by the same function $\overrightarrow{\mathrm{d}(TM)}$ (by the machine TM) with the same argument $[\overrightarrow{\mathrm{d}(TM)} * \overrightarrow{\mathrm{d}(TM)}]$ is now given. The Auto-Reference (56), (57) is then the *generative function* for the infinite sequences, nevertheless thought only, as the *consequence of the stationarity concept in-built in this type of consideration*,

$$\begin{gathered}
\left(TM, TM, ..., TM, ...\right) \triangleq \left(TM_{\Delta t_p}\right)_{p=1}^{\infty} \\
\left[\mathrm{d}(TM), \mathrm{d}(TM), ..., \mathrm{d}(TM), ...\right] \triangleq \left[\left[\mathrm{d}(TM)\right]_{\Delta t_p}\right]_{p=1}^{\infty} \\
\left[\boldsymbol{HALT}_{TM}^{\infty},\ \boldsymbol{HALT}_{TM}^{\infty}, ..., \boldsymbol{HALT}_{TM}^{\infty}, ...\right] \triangleq \left[\left(\boldsymbol{HALT}_{TM}^{\infty}\right)_{\Delta t_p}\right]_{p=1}^{\infty} \\
\left[\boldsymbol{LOOP}_{TM}^{\infty},\ \boldsymbol{LOOP}_{TM}^{\infty}, ..., \boldsymbol{LOOP}_{TM}^{\infty}, ...\right] \triangleq \left[\left(\boldsymbol{LOOP}_{TM}^{\infty}\right)_{\Delta t_p}\right]_{p=1}^{\infty}
\end{gathered} \tag{58}$$

We envisage, within this *time-expansion* of the (56) or (57) [which possibility follows from the (quasi)stationarity concept], the infinite cycle in the observed machine TM arises, based on its Self-Description $[\vec{\xi}] \equiv [\overrightarrow{\mathrm{d}(TM)}]$. But, following the Auto-Reference construction, it 'runs' in the double-machine $(TM, M_3 \equiv TM) \cong \mathcal{OO}''$.[12]

The Auto-Reference step that is to solve the Halting solve the Halting Problem proves, only, its own disusability **creates just a certain image of what is to be possibly discovered - the infinite cycle** in the form of the infinite constant time sequences [when the time expansion (58) for $p \geq 1$ is considered].

As it is valid for any ***stationary status****, also this one can be ceased or* ***can be excluded*** *by an external action, by the 'step-aside',* ***by the staging*** *as follows.*

5. Concept Removing Halting Problem

We suppose that in the case of a computing process running in a TM its status $\boldsymbol{LOOP}_{TM}^{\infty}$ (the infinite cycle) is decidable within the Observing Turing Machine (OTM) by using the $TM's$

12 Generally, the cycle $\mathcal{OO}''$ is considered as the reversible only.

trace. By this trace, the machine OTM generates and controls the 'combined observing process' for the process in the TM.

We will show and use the fact, that certain *regular sequences are generated. If they are infinite, they are, inevitably, periodical*; as such, they are *decidable languages for their infinity* [1].

We will use the alphabet of terminal symbols $\mathbf{T}=\{I,B\}$ and these structures:

- (s_i,x_k,s_j,y_l,D) is the *instruction*
- $(\vec{\sigma},s_{[.]},\vec{\rho})$ is the *configuration*
- $\left(\varepsilon\left[\sigma,s_{[.]},\rho\right]\varepsilon\right)$ is the *configuration type*

Further, we introduce the *general configuration type* $\boldsymbol{X}$;

- $\boldsymbol{X}=(\vec{\sigma},s_{[.]},\vec{\rho})\triangleq(\mathbf{B}\sigma,s_{[.]},\rho\,\mathbf{B})$.

By this general configuration type the chains, e.g. $\mathbf{BIB}s_{[.]}\mathbf{BIB}$ are ment,

$$\mathbf{BIB}s_{[\cdot]}\mathbf{BIB}=BBBBIIIIBBBBIIIIIBBBs_{[\cdot]}BBIIBBBIIIBBBB$$
$$\in\left(\vec{\sigma},s_{[\cdot]},\vec{\rho}\right)$$

The computing process in the observed TM generates the grammar of a regular language of instructions and, also, of general types of configurations especially, infinite possibly, and thus cyclical. As such, they are decidable languages for their infinity.

These grammars are given by the initial input $\vec{\xi}$, or, respectively, by the initial configuration $(\varepsilon,s_0,\vec{\xi})$ and, also, by the instructions η_{q_p} of the programme $\vec{\eta}$, being generated by the sequence of the steps p of the TM 's activity in which the TM instructions are interpreted. This sequence itself is pressed out by the configurations having been generated (See Appendix).

The Auto-Reference arises when, e.g., for a certain $p\geq p^*\geq p^0\geq 1$,

$$H\left(\overrightarrow{X_p}\right)-H\left(\overrightarrow{X_p}\mid\overrightarrow{X_{p+1}}\right)=H\left(\overrightarrow{X_{p+1}}\right)-H\left(\overrightarrow{X_{p+1}}\mid\overrightarrow{X_p}\right)=0$$
$$\overrightarrow{X_p}\triangleq\left(X_{p^*},X_{p^*+1},\ldots,X_p\right) \tag{59}$$
$$\overrightarrow{X_{p+1}}\triangleq\left(\overrightarrow{X_p},X_{p+1},\ldots,X_{2p+2-p^*}\right)\equiv\left(\overrightarrow{X_p}\overrightarrow{X_p}\right)$$

Also we can write [the similar is writable for (22), (23); also see the remark 7].

$$\begin{aligned}
&\overrightarrow{X_p}-\left(\overrightarrow{X_p}\sqsubseteq\overrightarrow{X_{p+1}}\right)=\overrightarrow{X_{p+1}}-\left(\overrightarrow{X_p}\sqsubseteq\overrightarrow{X_{p+1}}^{-1}\right)=\varepsilon\\
&\overrightarrow{X_p}=\left(\overrightarrow{X_p}\sqsubseteq\overrightarrow{X_{p+1}}\right)=\left(\overrightarrow{X_p}\sqsubseteq\overrightarrow{X_{p+1}}^{-1}\right)\\
&H\left(\overrightarrow{X_p}\right)=\left(\overrightarrow{X_p}\sqsubseteq\overrightarrow{X_{p+1}}\right)=H\left(\overrightarrow{X_p}\sqsubseteq\overrightarrow{X_{p+1}}^{-1}\right)=H\left(\overrightarrow{X_{p+1}}\sqsubseteq\overrightarrow{X_p}\right)=0\\
&\left[H\left(\overrightarrow{X_{p+1}}\right)=0\right]\Leftrightarrow\left[Pr\left(\overrightarrow{X_p}\right)=Pr\left(\overrightarrow{X_p}\overrightarrow{X_p}\right)=Pr\left(\left[\overrightarrow{X_p}\right]^{+}\right)=1\right]
\end{aligned}\tag{60}$$

where $Pr(\cdot)$ denotes *probability*.

5.1. Method - the *OTM*

✠^{1} Let the Turing Machine be driven by the program $\vec{\eta}$, do a certain number $e\cdot P$ of instructions $\eta_{[\cdot]}$ (of the program $\vec{\eta}$) beginning from the initial configuration $(\varepsilon,s_0,\vec{\xi})$. Let, e.g., $P=2l+1, l=|\vec{\xi}, e\geq 1$; $e\in\mathrm{N}$ be the number of the *stage* (for each stage we write, in a **programmer style**, $e:=e+1$)

- The step ✠^{1} generates the table of nine-partite structures. Its length of $e\cdot P$ rows,[13]

$$\left[p\,\|\,q\,\|\,s_i\,x_k\,s_j\,y_l\,D\,\|\left[\vec{\sigma}\,s_{[\cdot]}\,\vec{\rho}\right]\|\,C\,\|\|\left(\varepsilon\left[\boldsymbol{\sigma},s_{[\cdot]},\boldsymbol{\rho}\right]\varepsilon\right)\|\,g\,\|\,\vec{\boldsymbol{\sigma}}\,s_{[\cdot]}\,\vec{\boldsymbol{\rho}}\,\|\,m\right]_{p=1}^{p=e.P}\tag{61}$$

where the denotation used is

- C is the number of the *configuration* $(\vec{\sigma},s_{[\cdot]},\vec{\rho})$
- g is the number of the *configuration type* $(\varepsilon(\boldsymbol{\sigma},s_{[\cdot]},\boldsymbol{\rho})\varepsilon)$
- m is the number of the *general configuration type G*, $(\vec{\sigma},s_{[\cdot]},\vec{\rho})$

✠^{2} In the table (✠^{1}) we are seeking two successive blocks of rows limited by those rows having the identical values in the columns (the identical six-partite structures)

$$\left[q\,\|\,s_i\,x_k\,s_j\,y_l\,D\,\|\left(\varepsilon\left[\boldsymbol{\sigma},s_{[\cdot]},\boldsymbol{\rho}\right]\varepsilon\right)\|\,g\,\|\,\vec{\boldsymbol{\sigma}}\,s_{[\cdot]}\,\vec{\boldsymbol{\rho}}\,\|\,m\right]\tag{62}$$

Thus, we are seeking for (the sequence of) the *three rows* being identical in those columns while the last row of the first block is the first row of the second block [this second ends by the third row (identical in the six columns considered)]. The numbers of these *separating* rows, the first,

13 The symbols '[' and ']' in the tables denote the range of the input (its limits) and, also, the 'operating space' for the CU_{TM} in each step.

the second and the third row are the numbers of steps $p_{[\cdot]}, p_{[\cdot\cdot]}$ and $p_{[\cdots]}$ of the observed computing process. These rows are separated by numbers

$$\Delta p_{[\cdot]} \triangleq p_{[\cdot\cdot]} - p_{[\cdot]} - 1 \geq 0, \qquad \Delta p_{[\cdot\cdot]} \triangleq p_{[\cdots]} - p_{[\cdot\cdot]} - 1 \geq 0 \tag{63}$$

of rows lying between them. (They can follow immediately, $\Delta p_{[\cdot]}=0, \Delta p_{[\cdot\cdot]}=0$.)

Φ^3 If the three separating rows are **not found** within the given stage e (Φ^2), we start the computing process driven by the program $\vec{\eta}$, and its tracing, from the beginning [$(\varepsilon, s_0, \vec{\xi})$, ($\Phi^1$)], and let it run $e \cdot P$ steps, where $e := e+1$ is set down.

Φ^4 If those three separating rows are **found** within the given stage e (Φ^2) (those two blocks covering the rows $p_{[\cdot]}, p_{[\cdots]}$ where $p_{[\cdots]} = p_{[\cdot\cdot]} + \Delta p_{[\cdot\cdot]} + 1$ and $p_{[\cdot\cdot]} = p_{[\cdot]} + \Delta p_{[\cdot]} + 1$) we are checking both the two blocks, each of them from its beginning ($p_{[\cdot]}$, or $p_{[\cdot\cdot]}$ respectively) till its end ($p_{[\cdot\cdot]}$, or $p_{[\cdots]}$ respectively), seeking the rows with the identical values within their six columns (62),[14]

$$\left[q \,\|\, s_i \, x_k \, s_j \, y_l \, D \,\|\, \left(\varepsilon \left[\boldsymbol{\sigma}, s_{[\cdot]}, \boldsymbol{\rho} \right] \varepsilon \right) \|\, g \,\|\, \vec{\boldsymbol{\sigma}} s_{[\cdot]} \, \vec{\boldsymbol{\rho}} \,\|\, m \right]$$

Φ^5 Two or more such **identical six-partite structures** on the successive row positions (denoted by symbols z and m) are **considered as the only one row**

$$\begin{aligned}
&1) \quad z_{[\cdot]}, \ldots, z_{[\cdot\cdot]},\ p_{[\cdot]} = z_{[\cdot]} \leq z_{[\cdot\cdot]},\ m = 1 \overset{\text{Def}}{=} m_{[\cdot]} \\
&2) \quad z'_{[\cdot]}, \ldots, z'_{[\cdot\cdot]},\ z'_{[\cdot]} > z_{[\cdot\cdot]},\ z'_{[\cdot\cdot]} \geq z'_{[\cdot]},\ m = 2 \\
&3) \quad z''_{[\cdot]}, \ldots, z''_{[\cdot\cdot]},\ z''_{[\cdot]} > z_{[\cdot\cdot]},\ z''_{[\cdot\cdot]} \geq z''_{[\cdot]},\ m = 3 \\
&\qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&\cdot\ldots\cdot) \quad z^{\cdot\ldots\cdot}_{[\cdot]}, \ldots, z^{\cdot\ldots\cdot}_{[\cdot\cdot]},\ p_{[\cdot\cdot]} = z^{\cdot\ldots\cdot}_{[\cdot\cdot]} \geq z^{\cdot\ldots\cdot}_{[\cdot]},\ m = \cdot\ldots\cdot \overset{\text{Def}}{=} m_{[\cdot\cdot]}
\end{aligned} \tag{64}$$

(the first of them is considered only).[15]

Φ^6 We check whether, by this way, two new **identical successive blocks** of unique six-partite structures are created [the first block is between the (newly numbered) rows $m_{[\cdot]} \div m_{[\cdot\cdot]}$ and the second is (newly) between the rows $m_{[\cdot\cdot]} \div m_{[\cdots]}$].

Their lengths are $\Delta m_{[\cdot]}$ and $\Delta m_{[\cdot\cdot]}$ [for $\Delta m_{[\cdot\cdot]}$: $m_{[\cdot\cdot]} := (m_{[\cdot\cdot]})_{\mathrm{mod}(m_{[\cdot\cdot]}-1)}$],

14 If these blocks are identical the infinite cycle is discovered, but we continue uniformly with Φ^5.

15 The situation in the first block is described only by (64).

$$\Delta m_{[\cdot]} = m_{[\cdot\cdot]} - m_{[\cdot]} + 1 \le \Delta p_{[\cdot]},\ \Delta m_{[\cdot\cdot]} = m_{[\cdot\cdot\cdot]} - m_{[\cdot\cdot]} + 1 \le \Delta p_{[\cdot\cdot]}. \tag{65}$$

If NOT - we continue with Φ^2, $p^* = p_{[\ldots]}$; Φ^1 when $e \cdot P$ is exhausted, $e := e + 1$;

If YES **- the distances ($\Delta m_{[\cdot]}[-2]$ and $\Delta m_{[\cdot\cdot]}[-2]$) between the marginal rows of the new blocks ($\boxtimes^5$) are constant**, $\Delta m_{[\cdot]} = \Delta m_{[\cdot\cdot]}$,

[the distances are counted in the number m of the unique six-partite structures (Φ^5), the last one of the first block is the first one of the second block], $\Delta m_{[\cdot]} = \Delta m_{[\cdot\cdot]}$.

Φ^7 **If the distance $\Delta p_{[\cdot]}$ of the first and the last row of the first block ($\boxtimes^2$) is *less* or *equal* to the distance $\Delta p_{[\cdot\cdot]}$ of the first and last row of the second blok $\Delta p_{[\cdot]} \le \Delta p_{[\cdot\cdot]}$, we have discovered the infinite cycle driven by the program $\vec{\eta}$** [now the distances are counted in the number of steps p].

We continue further but **within the first block and with the $\Delta m_{[\cdot]}$ only.**

From each unique six-partite structures (Φ^5, Φ^6) the three columns

$$\left[q \| G \| m \right] \tag{66}$$

are now taken only, being interpreted as the rules of a regular grammar (accepted by a finite automaton)

$$\begin{gathered} S_{q^*} \to G_m S_{q_m} \\ q^* \in \{0\} \dot{\cup} \left\{1, \ldots, |\vec{\eta}|\right\}, \quad q_m \in \left\{1, \ldots, |\vec{\eta}|\right\}, \quad m = m_{[\cdot]}, \ldots, m_{[\cdot\cdot]} - 1 \end{gathered} \tag{67}$$

with the set $\mathbf{S}$ of nonterminal symbols and the set $\mathbf{T}'$ of terminal symbols,

$$\mathbf{S}_* \overset{\text{Def}}{=} \{S_0\} \dot{\cup} \{S_{q_m}\}_{m=m_{[\cdot]}}^{m_{[\cdot\cdot]}-1}, \ \operatorname{card} \mathbf{S}_* = n, \ \mathbf{T}' \overset{\text{Def}}{=} \{G_{q_m}\}_{m=m_{[\cdot]}}^{m_{[\cdot\cdot]}-1} \tag{68}$$

having the starting nonterminal symbol $S_0 \in \mathbf{S}$.[16]

16 Φ^{7a} If the distance $\Delta p_{[\cdot]}$ of the first and the last row of the first block (Φ^5 is greater than the distance $\Delta p_{[\cdot\cdot]}$ of the first and last row of the second blok, $\Delta p_{[\cdot]} > \Delta p_{[\cdot\cdot]}$, we have discovered the finite cycle driven by the program $\vec{\eta}$.

$\boxplus^{8}$ For the first block of the unique six–partite structures ($\boxtimes^{5}$, $\boxtimes^{6}$, $\boxtimes^{7}$) the sequence of rules of a regular grammar is being generated (accepted by the finite automaton with the states $S_0, S_{q_{m_{[.]}}}, \dots$)

$$\begin{aligned} S_0 &\rightarrow G_{m_{[.]}} S_{q_{m_{[.]}}} \\ S_{q_{m_{[.]}}} &\rightarrow G_{m_{[.]}+1} S_{q_{m_{[.]}+1}} \\ \dots &\quad \dots \quad \dots \\ S_{q_{m_{[..]}-2}} &\rightarrow G_{m_{[..]}-1} S_{q_{m_{[..]}-1}} \\ S_{q_{m_{[..]}-2}} &\rightarrow \vec{\varepsilon}\, S_0, \left[\text{or } S_{q_{m_{[..]}-1}} \rightarrow \varepsilon\, S_{HALT},\ S_{HALT} \notin \mathbf{S} \right] \end{aligned} \tag{69}$$

where the denotation $S_{q_{m_{[.]}}} \triangleq S_{z_{[.]},\dots,z_{[..]}}, S_{q_{m_{[.]}+1}} \triangleq S_{z_{[.]}',\dots,z_{[..]}'}, \dots S_{q_{m_{[..]}}} \triangleq S_{z_{[.]}'',\dots,z_{[..]}''}$ is used.

We have described the activity of a ***finite automaton*** *which accepts the infinite regular language of the general configurations types* (of the configurations of the observed machine *TM*). They are the words of the infinite length and having the form

$$\left[\left(\vec{\boldsymbol{\sigma}}_m, s_m, \vec{\boldsymbol{\rho}}_m \right)_{m=m_{[.]}}^{m_{[..]}-1} \right]^{+} = \left[\overrightarrow{\boldsymbol{X}} \right]^{+} \quad \left[\triangleq \left[\mathbf{d}(TM^{*}) \right]^{+} \right] \tag{70}$$

Yet after the *second round* of the observed *TM* through the infinite cycle has been finished the Pumping Lemma is usable and valid for the length L of the relevant word of this infinite language, [card$\mathbf{S} \leq L < 2 \cdot$ card$\mathbf{S}$].

We can generate the status (the signal) S_{HALT} to halt the whole machine *OTM* and the *TM* consequently.

We can say, briefly, that: If such the three identical bi-partite structures $[\eta, (\vec{\sigma}, s, \vec{\rho})]$, following each other, exist that for their distances $\Delta p_{[.]}$ (measured by the number of steps of the observed process) between the first and the second bi-partite structure and $\Delta p_{[..]}$ between the second and the third bi-partite structure it is valid that $\Delta p_{[.]} \leq \Delta p_{[..]}$, the observed *TM* is going through the infinite cycle.

The expression (70) means that we have discovered, within the dynamical system $\mathcal{O}\mathcal{O}''$, the dynamical subsystem $\mathcal{O}^{*}\mathcal{O}^{*'}$ ($\equiv TM^{*}$) which is in a *limit cycle*. It means the thermodynamic equilibrium within the double cycle $\mathcal{O}^{*}\mathcal{O}^{*'}$, thus for its temperatures it is valid that $T^{*}_{W} = T^{*'}_{W}$,

$T^*_0 = T^{*'}_0$; for sequences of the general configuration types ($G_{[\cdot]}$), $\overrightarrow{\boldsymbol{X}}_p$ and $\overrightarrow{\boldsymbol{X}}_{p+1}$ from (59), (60), it is valid, for certain $p \geq p^* \geq p^0 \geq 1$, that

$$\begin{aligned} &H\left(\overrightarrow{\boldsymbol{X}}_p\right) - H\left(\overrightarrow{\boldsymbol{X}}_p \mid \overrightarrow{\boldsymbol{X}}_{p+1}\right) = 0 \ \text{ where} \\ &\overrightarrow{\boldsymbol{X}}_p = \left(\boldsymbol{X}_{p+1}, ..., \boldsymbol{X}_{2p+2-p^*}\right) = \left(\boldsymbol{X}_{p^*}, ..., \boldsymbol{X}_p\right) \end{aligned} \tag{71}$$

which represents the *zero change of the information* and *thermodynamic entropy* within the working medium of a reversible Carnot Cycle; for the sequences $\overrightarrow{X}_p$ and $\overrightarrow{X}_{p+1}$ of the observed TM 's configurations $X_{[\cdot]}$ from (47) ($TM \equiv O$) it is valid that

$$H(\overrightarrow{X}_p) - H\left(\overrightarrow{X}_p \mid \overrightarrow{X}_{p+1}\right) = H(\overrightarrow{Y}_p) > 0, \ \overrightarrow{X}_{[\cdot]} \triangleq X_{p_*}, ..., X_{[\cdot]}, \ p^* \geq 1 \tag{72}$$

which represents the *nonzero output and, also, the growth of the thermodynamic and information entropy* within the whole isolated system in which that reversible Carnot Cycle is running, see [6, 8]. Generally, any Carnot Cycle is, under its construction draft, the infinite cycle, and, thus, both the relations (71) and (72) represent the *information thermodynamic criterion for the infinite cycle* existence.

The following section gives the examples of this method.

6. Examples

Example I

$$\begin{aligned} &\vec{\eta} = \left(\eta_1, \ \eta_2, \ \eta_3, \ \eta_4\right); \\ &\eta_1 = \left(s_0, \ I, \ s_0, \ I, \ R\right) \\ &\eta_2 = \left(s_0, \ B, \ s_1, \ I, \ L\right) \\ &\eta_3 = \left(s_1, \ I, \ s_1, \ I, \ L\right) \\ &\eta_4 = \left(s_1, \ B, \ s_0, \ B, \ R\right) \end{aligned} \tag{73}$$

This program *conserves* the given string $\vec{\xi}$

$$\vec{\xi} = \left[IIII...I\right] \ \text{resp.} \ B\left[IIII...I\right]B \ \text{resp.} \ B\left[IIII...I\right]B \tag{74}$$

or **BIB** respectively. Let the input $\vec{\xi} = IIII$ be given. From the table Tab. 1 it is obvious that

$$\begin{aligned} &p_{[\cdot]} = 1,\ p_{[\cdot\cdot]} = 13,\ p_{[\cdots]} = 25,\ m_{[\cdot]} = 1,\ m_{[\cdot\cdot]} = 7 \ \text{ and thus} \\ &\Delta p_{[\cdot]} = 11,\ \Delta p_{[\cdot\cdot]} = 13,\ \Delta p_{[\cdots]} = 15,\ \Delta m_{[\cdot]} = \Delta m_{[\cdot\cdot]} = \Delta m_{[\cdots]} = \ldots = 7 \end{aligned} \tag{75}$$

Then we can write down the regular grammar of the (regular) language of the general configuration given by the computing process driven by the program $\vec{\eta}$,

$$\begin{aligned} &S_0 \to \vec{\boldsymbol{B}}\, s_0\, \overrightarrow{\boldsymbol{IB}}\, S_1 \\ &S_1 \to \overrightarrow{\boldsymbol{BI}}\, s_0\, \overrightarrow{\boldsymbol{IB}}\, S_{2345} \\ &S_{2345} \to \overrightarrow{\boldsymbol{BI}}\, s_0\, \vec{\boldsymbol{B}}\, S_6 \\ &S_6 \to \overrightarrow{\boldsymbol{BI}}\, s_1\, \overrightarrow{\boldsymbol{IB}}\, S_{78910} \\ &S_{78910} \to \vec{\boldsymbol{B}}\, s_1\, \overrightarrow{\boldsymbol{IB}}\, S_{11} \\ &S_{11} \to \vec{\boldsymbol{B}}\, s_1\, \overrightarrow{\boldsymbol{BIB}}\, S_{12} \\ &S_{12} \to \vec{\varepsilon}\, S_0 \end{aligned} \tag{76}$$

This grammar is with the set $\mathbf{S}_*$ of its nonterminal symbols,

$$\mathbf{S}_* = \left\{S_0\ \ S_1\ \ S_{2345}\ \ S_6\ \ S_{78910}\ \ S_{11}\, S_{12}\right\},\ \text{card}\, \mathbf{S}_* = 7 \tag{77}$$

(let us remember (66)-(69) and that $S_1 \triangleq S_{q_1}$, $S_2 \triangleq S_{q_2}$, $S_6 \triangleq S_{q_3}$, $S_7 \triangleq S_{q_4}$, $S_{11} \triangleq S_{q_5}$, $S_{12}=S_{q_6}$)

and generates (the computing process described by it generates) the infinite word $[\,\overrightarrow{X}\,]^+$,

$$\left[\overrightarrow{\boldsymbol{X}}\right]^+ = \left[\vec{\boldsymbol{B}}\, s_0\, \overrightarrow{\boldsymbol{IB}}, \overrightarrow{\boldsymbol{BI}}\, s_0\, \overrightarrow{\boldsymbol{IB}}, \overrightarrow{\boldsymbol{BI}}\, s_0\, \vec{\boldsymbol{B}}, \overrightarrow{\boldsymbol{BI}}\, s_1\, \overrightarrow{\boldsymbol{IB}}, \vec{\boldsymbol{B}}\, s_1\, \overrightarrow{\boldsymbol{IB}}, \vec{\boldsymbol{B}}\, s_1\, \overrightarrow{\boldsymbol{BIB}}\right]^+ = \left[\mathbf{d}(TM^*)\right]^+ \tag{78}$$

For the generation of the signal which halts the whole combined machine we can add and use the rule

$$S_{12} \to \varepsilon\, S_{HALT}\,,\ S_{HALT} \notin \mathbf{S}_* \tag{79}$$

After the second round through the indicated infinite cycle the word of the general configuration types of the length l=12 is generated out and, thus, the Pumping Lemma it is valid,

$$\begin{aligned} &\text{card}\, \mathbf{S}_* \le l < 2 \cdot (\text{card}\, \mathbf{S}_*), \qquad \text{card}\, \mathbf{S}_* = 7 \\ &7 \le 12 < 2 \cdot 7 \end{aligned} \tag{80}$$

See the following table.

p ♯♯♯♯	q ♯♯♯♯	**Instruction**	**Configuration**	C ♯♯♯♯	**Config. Type**	g ♯♯♯♯	**Gerneral Config. Type** G	m ♯♯♯♯
1	**1**	$s_0\ I\ s_0\ I\ R$	$\varepsilon B[\ s_0\ IIIII]B\varepsilon$	**1**	$\varepsilon\boldsymbol{B}[\ s_0\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	**1**	$\overrightarrow{\boldsymbol{B}}\ s_0\ \overrightarrow{\boldsymbol{IB}}$	**1**
2	1	$s_0\ I\ s_0\ I\ R$	$\varepsilon B[I\ s_0\ IIII]B\varepsilon$	2	$\varepsilon\boldsymbol{B}[\boldsymbol{I}\ s_0\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	2	$\overrightarrow{\boldsymbol{BI}}\ s_0\ \overrightarrow{\boldsymbol{IB}}$	2
3	1	$s_0\ I\ s_0\ I\ R$	$\varepsilon B[II\ s_0\ III]B\varepsilon$	3	$\varepsilon\boldsymbol{B}[\boldsymbol{I}\ s_0\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	2	.....	2
4	1	$s_0\ I\ s_0\ I\ R$	$\varepsilon B[III\ s_0\ II]B\varepsilon$	4	$\varepsilon\boldsymbol{B}[\boldsymbol{I}\ s_0\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	2	.....	2
5	1	$s_0\ I\ s_0\ I\ R$	$\varepsilon B[IIII\ s_0\ I]B\varepsilon$	5	$\varepsilon\boldsymbol{B}[\boldsymbol{I}\ s_0\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	2	.....	2
6	2	$s_0\ B\ s_1\ B\ L$	$\varepsilon B[IIIII\ s_0\ ]B\varepsilon$	6	$\varepsilon\boldsymbol{B}[\boldsymbol{I}\ s_0\ ]\boldsymbol{B}\varepsilon$	3	$\overrightarrow{\boldsymbol{BI}}\ s_0\ \boldsymbol{B}$	3
7	3	$s_1\ I\ s_1\ I\ L$	$\varepsilon B[IIII\ s_1\ IB]\varepsilon$	7	$\varepsilon\boldsymbol{B}[\boldsymbol{I}\ s_1\ \boldsymbol{IB}]\varepsilon$	4	$\overrightarrow{\boldsymbol{BI}}\ s_1\ \overrightarrow{\boldsymbol{IB}}$	4
8	3	$s_1\ I\ s_1\ I\ L$	$\varepsilon B[III\ s_1\ IIB]\varepsilon$	8	$\varepsilon\boldsymbol{B}[\boldsymbol{I}\ s_1\ \boldsymbol{IB}]\varepsilon$	4	.....	4
9	3	$s_1\ I\ s_1\ I\ L$	$\varepsilon B[II\ s_1\ IIIB]\varepsilon$	9	$\varepsilon\boldsymbol{B}[\boldsymbol{I}\ s_1\ \boldsymbol{IB}]\varepsilon$	4	.....	4
10	3	$s_1\ I\ s_1\ I\ L$	$\varepsilon B[I\ s_1\ IIIIB]\varepsilon$	10	$\varepsilon\boldsymbol{B}[\boldsymbol{I}\ s_1\ \boldsymbol{IB}]\varepsilon$	4	.....	4
11	3	$s_1\ I\ s_1\ I\ L$	$\varepsilon B[\ s_1\ IIIIIB]\varepsilon$	11	$\varepsilon\boldsymbol{B}[\ s_1\ \boldsymbol{IB}]\varepsilon$	5	$\overrightarrow{\boldsymbol{B}}\ s_1\ \overrightarrow{\boldsymbol{IB}}$	5
12	4	$s_1\ B\ s_0\ B\ R$	$\varepsilon[\ s_1\ BIIIIIB]\varepsilon$	12	$\varepsilon[\ s_1\ \boldsymbol{BIB}]\varepsilon$	6	$\overrightarrow{\boldsymbol{B}}\ s_1\ \overrightarrow{\boldsymbol{BIB}}$	6
13	**1**	$s_0\ I\ s_0\ I\ R$	$\varepsilon[B\ s_0\ IIIIIB]\varepsilon$	**1′**	$\varepsilon[\boldsymbol{B}\ s_0\ \boldsymbol{IB}]\varepsilon$	**1**	$\overrightarrow{\boldsymbol{B}}\ s_0\ \overrightarrow{\boldsymbol{IB}}$	**7**, **1**
14	1	$s_0\ I\ s_0\ I\ R$	$\varepsilon[BI\ s_0\ IIIIB]\varepsilon$	2′	$\varepsilon[\boldsymbol{BI}\ s_0\ \boldsymbol{IB}]\varepsilon$	2	$\overrightarrow{\boldsymbol{BI}}\ s_0\ \overrightarrow{\boldsymbol{IB}}$	2
15	1	$s_0\ I\ s_0\ I\ R$	$\varepsilon[BII\ s_0\ IIIB]\varepsilon$	3′	$\varepsilon[\boldsymbol{BI}\ s_0\ \boldsymbol{IB}]\varepsilon$	2	.....	2
16	1	$s_0\ I\ s_0\ I\ R$	$\varepsilon[BIII\ s_0\ IIB]\varepsilon$	4′	$\varepsilon[\boldsymbol{BI}\ s_0\ \boldsymbol{IB}]\varepsilon$	2	.....	2
17	1	$s_0\ I\ s_0\ I\ R$	$\varepsilon[BIIII\ s_0\ IB]\varepsilon$	5′	$\varepsilon[\boldsymbol{BI}\ s_0\ \boldsymbol{IB}]\varepsilon$	2	.....	2
18	2	$s_0\ B\ s_1\ B\ L$	$\varepsilon[BIIIII\ s_0\ B]\varepsilon$	6′	$\varepsilon[\boldsymbol{BI}\ s_0\ \boldsymbol{B}]\varepsilon$	3	$\overrightarrow{\boldsymbol{BI}}\ s_0\ \overrightarrow{\boldsymbol{B}}$	3
19	3	$s_1\ I\ s_1\ I\ L$	$\varepsilon[BIIII\ s_1\ IB]\varepsilon$	7′	$\varepsilon[\boldsymbol{BI}\ s_1\ \boldsymbol{IB}]\varepsilon$	4	$\overrightarrow{\boldsymbol{BI}}\ s_1\ \overrightarrow{\boldsymbol{IB}}$	4
20	3	$s_1\ I\ s_1\ I\ L$	$\varepsilon[BIII\ s_1\ IIB]\varepsilon$	8′	$\varepsilon[\boldsymbol{BI}\ s_1\ \boldsymbol{IB}]\varepsilon$	4	.....	4
21	3	$s_1\ I\ s_1\ I\ L$	$\varepsilon[BII\ s_1\ IIIB]\varepsilon$	9′	$\varepsilon[\boldsymbol{BI}\ s_1\ \boldsymbol{IB}]\varepsilon$	4	.....	4
22	3	$s_1\ I\ s_1\ I\ L$	$\varepsilon[BI\ s_1\ IIIIB]\varepsilon$	10′	$\varepsilon[\boldsymbol{BI}\ s_1\ \boldsymbol{IB}]\varepsilon$	4	.....	4
23	3	$s_1\ I\ s_1\ I\ L$	$\varepsilon[B\ s_1\ IIIIIB]\varepsilon$	11′	$\varepsilon[\boldsymbol{B}\ s_1\ \boldsymbol{IB}]\varepsilon$	5	$\overrightarrow{\boldsymbol{B}}\ s_1\ \overrightarrow{\boldsymbol{IB}}$	5
24	4	$s_1\ B\ s_0\ B\ R$	$\varepsilon[\ s_1\ BIIIIIB]\varepsilon$	12′	$\varepsilon[\boldsymbol{B}\ s_1\ \boldsymbol{BIB}]\varepsilon$	6	$\overrightarrow{\boldsymbol{B}}\ s_1\ \overrightarrow{\boldsymbol{BIB}}$	6
25	**1**	$s_0\ I\ s_0\ I\ R$	$\varepsilon[B\ s_0\ IIIIIB]\varepsilon$	**1″**	$\varepsilon[\boldsymbol{B}\ s_0\ \boldsymbol{IB}]\varepsilon$	**1**	$\overrightarrow{\boldsymbol{B}}\ s_0\ \overrightarrow{\boldsymbol{IB}}$	**7**, **1**
26	..	.....	....	2″	....	2	.....	2
..	..	.....	....	..	....	..	.....	..
36	..	.....	....	12″	....	6	.....	6
37	**1**	$s_0\ I\ s_0\ I\ R$	$\varepsilon[B\ s_0\ IIIIIB]\varepsilon$	**1‴**	$\varepsilon[\boldsymbol{B}\ s_0\ \boldsymbol{IB}]\varepsilon$	**1**	$\overrightarrow{\boldsymbol{B}}\ s_0\ \overrightarrow{\boldsymbol{IB}}$	**1**
38	..	.....	....	2‴	....	2	.....	2
..	..	.....	....	..	....	..	.....	..
48	..	.....	....	12‴	....	6	.....	6
49	**1**	$s_0\ I\ s_0\ I\ R$	$\varepsilon[B\ s_0\ IIIIIB]\varepsilon$	**1⁗**	$\varepsilon[\boldsymbol{B}\ s_0\ \boldsymbol{IB}]\varepsilon$	**1**	$\overrightarrow{\boldsymbol{B}}\ s_0\ \overrightarrow{\boldsymbol{IB}}$	**1**

Table 1. Tracing and staging for Example I

[We 'idle' for the same number of clicks Δp in each of the blocks Δm. The sequence of Δp is (11, 11, 11, …), the sequence of Δm is (5, 5, 5, …). The sequence for p is (13, 13, 13, …) and for m is (7, 7, 7, …).]

After the *observed (sub)machine* has entered into the infinite cycle, which takes place in a finite time, and has gone through this cycle twice *it is halted by the signal from the observing machine*.

Example II

$$\begin{aligned}
\vec{\eta} &= \left(\eta_1,\ \eta_2,\ \eta_3,\ \eta_4\right)\\
\eta_1 &= \left(s_0,\ I,\ s_0,\ I,\ R\right)\\
\eta_2 &= \left(s_0,\ B,\ s_1,\ I,\ L\right)\\
\eta_3 &= \left(s_1,\ I,\ s_1,\ I,\ L\right)\\
\eta_4 &= \left(s_1,\ B,\ s_0,\ B,\ R\right)
\end{aligned} \tag{81}$$

This program *generates* the *expanding* sequence

$$\left[IIIII\right] \text{ resp. } B\left[IIIII...I\right]B,\ldots, B\left[IIII...I...\right]B,\ldots \tag{82}$$

or $\overset{\cdot}{\mathbf{BIB}}$ respectively. Let the input $\vec{\xi} = IIIII$ be given. From the following table Tab. 2 it is obvious that

$$\begin{aligned}
&p_{[\cdot]} = 1,\ p_{[\cdot\cdot]} = 13,\ p_{[\cdot\cdot\cdot]} = 27,\ m_{[\cdot]} = 1,\ m_{[\cdot\cdot]} = 7 \text{ and thus}\\
&\Delta p_{[\cdot]} = 11,\ \Delta p_{[\cdot\cdot]} = 13,\ \Delta p_{[\cdot\cdot\cdot]} = 15,\ \Delta m_{[\cdot]} = \Delta m_{[\cdot\cdot]} = \Delta m_{[\cdot\cdot\cdot]} = \ldots = 7
\end{aligned} \tag{83}$$

and we can write down the regular grammar (of the regular) language of the general configuration types generated by the computing process driven by $\vec{\eta}$

$$\begin{aligned}
S_0 &\to \vec{\boldsymbol{B}}\, s_0\, \overrightarrow{\boldsymbol{IB}}\, S_1\\
S_1 &\to \overrightarrow{\boldsymbol{BI}}\, s_0\, \overrightarrow{\boldsymbol{IB}}\, S_{2345}\\
S_{2345} &\to \overrightarrow{\boldsymbol{BI}}\, s_0\, \vec{\boldsymbol{B}}\, S_6\\
S_6 &\to \overrightarrow{\boldsymbol{BI}}\, s_1\, \overrightarrow{\boldsymbol{IB}}\, S_{78910}\\
S_{78910} &\to \vec{\boldsymbol{B}}\, s_1\, \overrightarrow{\boldsymbol{IB}}\, S_{11}\\
S_{11} &\to \vec{\boldsymbol{B}}\, s_1\, \overrightarrow{\boldsymbol{BIB}}\, S_{12}\\
S_{12} &\to \vec{\varepsilon}\, S_0
\end{aligned} \tag{84}$$

where, after the relations (66)-(69), $S_1 \triangleq S_{q_1}$, $S_2 \triangleq S_{q_2}$, $S_6 \triangleq S_{q_3}$, $S_7 \triangleq S_{q_4}$, $S_{11} \triangleq S_{q_5}$, $S_{12}=S_{q_6}$.

This grammar is with the set of nonterminal symbols

$$\mathbf{S}_* = \{S_0 \;\; S_1 \;\; S_{2345} \;\; S_6 \;\; S_{78910} \;\; S_{11}\, S_{12}\}, \;\; \text{card}\, \mathbf{S}_* = 7 \tag{85}$$

and generates (the computing process described by it generates) the infinite word $[\overrightarrow{X}]^+$,

$$\left[\overrightarrow{X}\right]^+ = \left[\vec{B}\, s_0\, \overrightarrow{IB}, \overrightarrow{BI}\, s_0\, \overrightarrow{IB}, \overrightarrow{BI}\, s_0\, \vec{B}, \overrightarrow{BI}\, s_1\, \overrightarrow{IB}, \vec{B}\, s_1\, \overrightarrow{IB}, \vec{B}\, s_1\, \overrightarrow{BIB}\right]^+ = \left[\mathbf{d}(TM^*)\right]^+ \tag{86}$$

For the generation of the signal which halts the whole combined machine we can add and use the rule

$$S_{12} \to \varepsilon\, S_{HALT}\,, \;\; S_{HALT} \notin \mathbf{S}_* \tag{87}$$

After the second round through the indicated infinite cycle the word of the general configuration types of the length $l=12$ is generated and, thus, the Pumping Lemma it is valid,

$$\begin{aligned} &\text{card}\, \mathbf{S}_* \le l < 2 \cdot (\text{card}\, \mathbf{S}_*), \qquad \text{card}\, \mathbf{S}_* = 7 \\ &7 \le 12 < 2 \cdot 7 \end{aligned} \tag{88}$$

See the following table.

We are going through the seven states repeatedly and each such a pass lasts longer than the previous one.

We 'idle' in such each pass a longer time (for the growing number of clicks p of the CU_{TM}).

[The growing sequence of Δp is (11, 13, 15,...) is evidenced while the sequence of Δm is (5,5,5,...).

The sequence for p is (13, 15, 17,...) and for m is (7, 7, 7,...).]

After the *observed (sub)machine* has entered into the infinite cycle, which occurs in a finite time, and has gone through this cycle twice *it is halted by the signal from the observing machine*.

p ♯♯♯♯	q ♯♯♯♯	**Instruction**	**Configuration**	C ♯♯♯♯	**Config. Type**	g ♯♯♯♯	**Gerneral Config. Type** G	m ♯♯♯♯
1	**1**	$s_0\ I\ s_0\ I\ R$	$\varepsilon B[s_0\ IIIII]B\varepsilon$	**1**	$\varepsilon\boldsymbol{B}[s_0\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	**1**	$\overrightarrow{\boldsymbol{B}}\ s_0\ \overrightarrow{\boldsymbol{IB}}$	**1**
2	1	$s_0\ I\ s_0\ I\ R$	$\varepsilon B[I\ s_0\ IIII]B\varepsilon$	2	$\varepsilon\boldsymbol{B}[\boldsymbol{I}\ s_0\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	2	$\overrightarrow{\boldsymbol{BI}}\ s_0\ \overrightarrow{\boldsymbol{IB}}$	**2**
3	1	$s_0\ I\ s_0\ I\ R$	$\varepsilon B[II\ s_0\ III]B\varepsilon$	3	$\varepsilon\boldsymbol{B}[\boldsymbol{I}\ s_0\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	2	.....	2
4	1	$s_0\ I\ s_0\ I\ R$	$\varepsilon B[III\ s_0\ II]B\varepsilon$	4	$\varepsilon\boldsymbol{B}[\boldsymbol{I}\ s_0\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	2	.....	2
5	1	$s_0\ I\ s_0\ I\ R$	$\varepsilon B[IIII\ s_0\ I]B\varepsilon$	5	$\varepsilon\boldsymbol{B}[\boldsymbol{I}\ s_0\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	2	.....	2
6	2	$s_0\ B\ s_1\ I\ L$	$\varepsilon B[IIIII\ s_0\ B]\varepsilon$	6	$\varepsilon\boldsymbol{B}[\boldsymbol{I}\ s_0\ \boldsymbol{B}]\varepsilon$	3	$\overrightarrow{\boldsymbol{BI}}\ s_0\ \overrightarrow{\boldsymbol{B}}$	**3**
7	3	$s_1\ I\ s_1\ I\ L$	$\varepsilon B[IIII\ s_1\ II]\varepsilon$	7	$\varepsilon\boldsymbol{B}[\boldsymbol{I}\ s_1\ \boldsymbol{I}]\varepsilon$	4	$\overrightarrow{\boldsymbol{BI}}\ s_1\ \overrightarrow{\boldsymbol{IB}}$	**4**
8	3	$s_1\ I\ s_1\ I\ L$	$\varepsilon B[III\ s_1\ III]\varepsilon$	8	$\varepsilon\boldsymbol{B}[\boldsymbol{I}\ s_1\ \boldsymbol{I}]\varepsilon$	4	.....	4
9	3	$s_1\ I\ s_1\ I\ L$	$\varepsilon B[II\ s_1\ IIII]\varepsilon$	9	$\varepsilon\boldsymbol{B}[\boldsymbol{I}\ s_1\ \boldsymbol{I}]\varepsilon$	4	.....	4
10	3	$s_1\ I\ s_1\ I\ L$	$\varepsilon B[I\ s_1\ IIIII]\varepsilon$	10	$\varepsilon\boldsymbol{B}[\boldsymbol{I}\ s_1\ \boldsymbol{I}]\varepsilon$	4	.....	4
11	3	$s_1\ I\ s_1\ I\ L$	$\varepsilon B[s_1\ IIIIII]\varepsilon$	11	$\varepsilon\boldsymbol{B}[s_1\ \boldsymbol{I}]\varepsilon$	5	$\overrightarrow{\boldsymbol{B}}\ s_1\ \overrightarrow{\boldsymbol{IB}}$	**5**
12	4	$s_1\ B\ s_0\ B\ R$	$\varepsilon[s_1\ BIIIIII]\varepsilon$	12	$\varepsilon[s_1\ \boldsymbol{BI}]\varepsilon$	6	$\overrightarrow{\boldsymbol{B}}\ s_1\ \overrightarrow{\boldsymbol{BIB}}$	**6**
13	**1**	$s_0\ I\ s_0\ I\ R$	$\varepsilon[B\ s_0\ IIIIII]B\varepsilon$	**1′**	$\varepsilon[\boldsymbol{B}\ s_0\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	**1**	$\overrightarrow{\boldsymbol{B}}\ s_0\ \overrightarrow{\boldsymbol{IB}}$	**7,1**
14	1	$s_0\ I\ s_0\ I\ R$	$\varepsilon[BI\ s_0\ IIIII]B\varepsilon$	2′	$\varepsilon[\boldsymbol{BI}\ s_0\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	2	$\overrightarrow{\boldsymbol{BI}}\ s_0\ \overrightarrow{\boldsymbol{IB}}$	**2**
15	1	$s_0\ I\ s_0\ I\ R$	$\varepsilon[BII\ s_0\ IIII]B\varepsilon$	3′	$\varepsilon[\boldsymbol{BI}\ s_0\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	2	.....	2
16	1	$s_0\ I\ s_0\ I\ R$	$\varepsilon[BIII\ s_0\ III]B\varepsilon$	4′	$\varepsilon[\boldsymbol{BI}\ s_0\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	2	.....	2
17	1	$s_0\ I\ s_0\ I\ R$	$\varepsilon[BIIII\ s_0\ II]B\varepsilon$	5′	$\varepsilon[\boldsymbol{BI}\ s_0\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	2	.....	2
18	1	$s_0\ I\ s_0\ I\ R$	$\varepsilon[BIIIII\ s_0\ I]B\varepsilon$	5′	$\varepsilon[\boldsymbol{BI}\ s_0\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	2	.....	2
19	2	$s_0\ B\ s_1\ I\ L$	$\varepsilon[BIIIIII\ s_0\ B]\varepsilon$	6′	$\varepsilon[\boldsymbol{BI}\ s_0\ \boldsymbol{B}]\varepsilon$	3	$\overrightarrow{\boldsymbol{BI}}\ s_0\ \overrightarrow{\boldsymbol{B}}$	**3**
20	3	$s_1\ I\ s_1\ I\ L$	$\varepsilon[BIIIII\ s_1\ II]\varepsilon$	7′	$\varepsilon[\boldsymbol{BI}\ s_1\ \boldsymbol{I}]\varepsilon$	4	$\overrightarrow{\boldsymbol{BI}}\ s_1\ \overrightarrow{\boldsymbol{IB}}$	**4**
21	3	$s_1\ I\ s_1\ I\ L$	$\varepsilon[BIIII\ s_1\ III]\varepsilon$	8′	$\varepsilon[\boldsymbol{BI}\ s_1\ \boldsymbol{I}]\varepsilon$	4	.....	4
22	3	$s_1\ I\ s_1\ I\ L$	$\varepsilon[BIII\ s_1\ IIII]\varepsilon$	9′	$\varepsilon[\boldsymbol{BI}\ s_1\ \boldsymbol{I}]\varepsilon$	4	.....	4
23	3	$s_1\ I\ s_1\ I\ L$	$\varepsilon[BII\ s_1\ IIIII]\varepsilon$	10′	$\varepsilon[\boldsymbol{BI}\ s_1\ \boldsymbol{I}]\varepsilon$	4	.....	4
24	3	$s_1\ I\ s_1\ I\ L$	$\varepsilon[BI\ s_1\ IIIIII]\varepsilon$	10′	$\varepsilon[\boldsymbol{BI}\ s_1\ \boldsymbol{I}]\varepsilon$	4	.....	4
25	3	$s_1\ I\ s_1\ I\ L$	$\varepsilon[B\ s_1\ IIIIIII]\varepsilon$	11′	$\varepsilon[\boldsymbol{B}\ s_1\ \boldsymbol{I}]\varepsilon$	5	$\overrightarrow{\boldsymbol{B}}\ s_1\ \overrightarrow{\boldsymbol{IB}}$	**5**
26	4	$s_1\ B\ s_0\ B\ R$	$\varepsilon[s_1\ BIIIIIII]\varepsilon$	12′	$\varepsilon[s_1\ \boldsymbol{BI}]\varepsilon$	6	$\overrightarrow{\boldsymbol{B}}\ s_1\ \overrightarrow{\boldsymbol{BIB}}$	**6**
27	**1**	$s_0\ I\ s_0\ I\ R$	$\varepsilon[B\ s_0\ IIIIIII]B\varepsilon$	**1″**	$\varepsilon[\boldsymbol{B}\ s_0\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	**1**	$\overrightarrow{\boldsymbol{B}}\ s_0\ \overrightarrow{\boldsymbol{IB}}$	**7,1**

p ♯♯♯♯	q ♯♯♯♯	**Instruction**	**Configuration**	C ♯♯♯♯	**Config. Type**	g ♯♯♯♯	**Gerneral Config. Type** G	m ♯♯♯♯
27	**1**	$s_0\ I\ s_0\ I\ R$	$\varepsilon[B\ s_0\ IIIIIII]B\varepsilon$	$\mathbf{1}''$	$\varepsilon[\boldsymbol{B}\ s_0\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	**1**	$\overrightarrow{\boldsymbol{B}}\ s_0\ \overrightarrow{\boldsymbol{IB}}$	**1**
28	1	$s_0\ I\ s_0\ I\ R$	$\varepsilon[BI\ s_0\ IIIIII]B\varepsilon$	$2''$	$\varepsilon[\boldsymbol{BI}\ s_0\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	2	$\overrightarrow{\boldsymbol{BI}}\ s_0\ \overrightarrow{\boldsymbol{IB}}$	**2**
29	1	$s_0\ I\ s_0\ I\ R$	$\varepsilon[BII\ s_0\ IIIII]B\varepsilon$	$3''$	$\varepsilon[\boldsymbol{BI}\ s_0\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	2	.....	2
30	1	$s_0\ I\ s_0\ I\ R$	$\varepsilon[BIII\ s_0\ IIII]B\varepsilon$	$4''$	$\varepsilon[\boldsymbol{BI}\ s_0\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	2	.....	2
31	1	$s_0\ I\ s_0\ I\ R$	$\varepsilon[BIIII\ s_0\ III]B\varepsilon$	$5''$	$\varepsilon[\boldsymbol{BI}\ s_0\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	2	.....	2
32	1	$s_0\ I\ s_0\ I\ R$	$\varepsilon[BIIIII\ s_0\ II]B\varepsilon$	$5''$	$\varepsilon[\boldsymbol{BI}\ s_0\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	2	.....	2
33	1	$s_0\ I\ s_0\ I\ R$	$\varepsilon[BIIIIII\ s_0\ I]B\varepsilon$	$5''$	$\varepsilon[\boldsymbol{BI}\ s_0\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	2	.....	2
34	2	$s_0\ B\ s_1\ I\ L$	$\varepsilon[BIIIIIII\ s_0\ B]\varepsilon$	$6''$	$\varepsilon[\boldsymbol{BI}\ s_0\ \boldsymbol{B}]\varepsilon$	3	$\overrightarrow{\boldsymbol{BI}}\ s_0\ \overrightarrow{\boldsymbol{B}}$	**3**
35	3	$s_1\ I\ s_1\ I\ L$	$\varepsilon[BIIIIII\ s_1\ II]\varepsilon$	$7''$	$\varepsilon[\boldsymbol{BI}\ s_1\ \boldsymbol{I}]\varepsilon$	4	$\overrightarrow{\boldsymbol{BI}}\ s_1\ \overrightarrow{\boldsymbol{IB}}$	**4**
36	3	$s_1\ I\ s_1\ I\ L$	$\varepsilon[BIIIII\ s_1\ III]\varepsilon$	$8''$	$\varepsilon[\boldsymbol{BI}\ s_1\ \boldsymbol{I}]\varepsilon$	4	.....	4
37	3	$s_1\ I\ s_1\ I\ L$	$\varepsilon[BIIII\ s_1\ IIII]\varepsilon$	$9''$	$\varepsilon[\boldsymbol{BI}\ s_1\ \boldsymbol{I}]\varepsilon$	4	.....	4
38	3	$s_1\ I\ s_1\ I\ L$	$\varepsilon[BIII\ s_1\ IIIII]\varepsilon$	$10''$	$\varepsilon[\boldsymbol{BI}\ s_1\ \boldsymbol{I}]\varepsilon$	4	.....	4
39	3	$s_1\ I\ s_1\ I\ L$	$\varepsilon[BII\ s_1\ IIIIII]\varepsilon$	$10''$	$\varepsilon[\boldsymbol{BI}\ s_1\ \boldsymbol{I}]\varepsilon$	4	.....	4
40	3	$s_1\ I\ s_1\ I\ L$	$\varepsilon[BI\ s_1\ IIIIIII]\varepsilon$	$10''$	$\varepsilon[\boldsymbol{BI}\ s_1\ \boldsymbol{I}]\varepsilon$	4	.....	4
41	3	$s_1\ I\ s_1\ I\ L$	$\varepsilon[B\ s_1\ IIIIIIII]\varepsilon$	$11''$	$\varepsilon[\boldsymbol{B}\ s_1\ \boldsymbol{I}]\varepsilon$	5	$\overrightarrow{\boldsymbol{B}}\ s_1\ \overrightarrow{\boldsymbol{IB}}$	**5**
42	4	$s_1\ B\ s_0\ B\ R$	$\varepsilon[s_1\ BIIIIIIII]\varepsilon$	$12''$	$\varepsilon[s_1\ \boldsymbol{BI}]\varepsilon$	6	$\overrightarrow{\boldsymbol{B}}\ s_1\ \overrightarrow{\boldsymbol{BIB}}$	**6**
43	**1**	$s_0\ I\ s_0\ I\ R$	$\varepsilon[B\ s_0\ IIIIIIII]B\varepsilon$	$\mathbf{1}'''$	$\varepsilon[\boldsymbol{B}\ s_0\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	**1**	$\overrightarrow{\boldsymbol{B}}\ s_0\ \overrightarrow{\boldsymbol{IB}}$	**7**, **1**
44	..	.....	....	$2'''$	....	2	.....	**2**
..	..	.....	....	..	....	..	.....	..
60	..	.....	....	$12'''$	....	6	.....	**6**
61	**1**	$s_0\ I\ s_0\ I\ R$	$\varepsilon[B\ s_0\ IIIIIIIII]B\varepsilon$	$\mathbf{1}''''$	$\varepsilon[\boldsymbol{B}\ s_0\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	**1**	$\overrightarrow{\boldsymbol{B}}\ s_0\ \overrightarrow{\boldsymbol{IB}}$	**7**, **1**
62	..	.....	....	$2''''$	....	2	.....	**2**
..	..	.....	....	..	....	..	.....	..
80	..	.....	....	$12''''$	....	6	.....	**6**
81	**1**	$s_0\ I\ s_0\ I\ R$	$\varepsilon[B\ s_0\ IIIIIIIII]B\varepsilon$	$\mathbf{1}'''''$	$\varepsilon[\boldsymbol{B}\ s_0\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	**1**	$\overrightarrow{\boldsymbol{B}}\ s_0\ \overrightarrow{\boldsymbol{IB}}$	**7**, **1**

Table 2. Tracing and staging for Example II

Example III

$$
\begin{aligned}
&\vec{\eta} = \left(\eta_1,\ \eta_2,\ \eta_3,\ \eta_4,\ \eta_5,\ \eta_6,\ \eta_7\ \ \eta_8,\ \eta_9,\ \eta_{10}\ \ \eta_{11},\ \eta_{12},\ \eta_{13}\right)\\
&\eta_1 = \left(s_0,\ I,\ s_1,\ B,\ R\right)\\
&\eta_2 = \left(s_1,\ I,\ s_1,\ I,\ R\right)\\
&\eta_3 = \left(s_1,\ B,\ s_2,\ B,\ R\right)\\
&\eta_4 = \left(s_2,\ I,\ s_2,\ I,\ R\right)\\
&\eta_5 = \left(s_2,\ B,\ s_3,\ B,\ L\right)\\
&\eta_6 = \left(s_3,\ I,\ s_4,\ B,\ L\right)\\
&\eta_7 = \left(s_4,\ B,\ s_H,\ B,\ R\right)\\
&\eta_8 = \left(s_4,\ I,\ s_5,\ I,\ L\right)\\
&\eta_9 = \left(s_5,\ I,\ s_5,\ I,\ L\right)\\
&\eta_{10} = \left(s_5,\ B,\ s_6,\ B,\ L\right)\\
&\eta_{11} = \left(s_6,\ I,\ s_6,\ I,\ L\right)\\
&\eta_{12} = \left(s_6,\ B,\ s_0,\ B,\ R\right)\\
&\eta_{13} = \left(s_0,\ B,\ s_0,\ B,\ L\right)
\end{aligned}
\tag{89}
$$

This program *generates*, or is figuring, the *difference* between the numbers 3 and 4 in this order,

$$3 - 4 \tag{90}$$

and *without checking the negativity* of the result. Thus it enters into the infinite cycle in which generates the left-direction expanding sequence

$$\left[...BBB...BIBBB\right] \text{ resp. } B\left[...BBB...BIBBB\right]B \triangleq \overrightarrow{\boldsymbol{BIB}} \tag{91}$$

Let now the input word $\vec{\xi} = IIIBIIII$ be given.

From the following table Tab 3. it is obvious that after the generation of the finite length's beginning part G, but of the infinite lengthy word of the general configurations numbered by values of m,

$$
\begin{array}{lllllllllllllllll}
\mathbf{G} = [& 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 \\
 & 1 & 2 & 3 & 5 & 6 & 7 & 8 & 10 & 12 & 13 & 15 & 16 & & & & \\
 & 1 & 17 & 18 & 6 & 8 & 19 & 20 & 21 &] & & & & & & &
\end{array}
\tag{92}
$$

the part $[22]^{+}$ follows. From the table it is obvious that the cycle is with the values

$$\begin{aligned} &p_{[\cdot]} = 41,\ p_{[\cdot\cdot]} = 42,\ p_{[\cdots]} = 43 \\ &m_{[\cdot]} = 41,\ m_{[\cdot\cdot]} = 42 \end{aligned} \tag{93}$$

We can write down the regular grammar (of the regular) language of the general configuration types generated by the computing process driven by $\vec{\eta}$

$$\begin{aligned} S_0 &\to \mathbf{G}\, S_{1\cdots 40} \\ S_{1\cdots 40} &\to \vec{\boldsymbol{B}}\, s_0\, \overrightarrow{\boldsymbol{BIB}}\, S_{41\,42\,43\cdots} \\ S_{41\,42\,43\cdots} &\to \vec{\boldsymbol{B}}\, s_0\, \overrightarrow{\boldsymbol{BIB}}\, S_{41\,42\,43\cdots} \end{aligned} \tag{94}$$

This grammar is with the set of nonterminal symbols

$$\mathbf{S}_* = \left\{ S_0\ \ S_{1\cdots 40}\ \ S_{41\,42\,43\cdots}\ \ S_6\ \ S_{7\,8\,9\,10}\ \ S_{11}\, S_{12} \right\},\ \text{card}\, \mathbf{S}_* = 3 \tag{95}$$

and generates the infinite word

$$\left[\mathbf{G} \left[22 \right]^{+} \right] \tag{96}$$

After the second round through the indicated infinite cycle the subword of the length $l = 12$ is generated out,

$$\left[\vec{\boldsymbol{B}}\, s_0\, \overrightarrow{\boldsymbol{BIB}}, \vec{\boldsymbol{B}}\, s_0\, \overrightarrow{\boldsymbol{BIB}}, \vec{\boldsymbol{B}}\, s_0\, \overrightarrow{\boldsymbol{BIB}} \right] \tag{97}$$

and thus, the Pumping Lemma it is valid,

$$\begin{aligned} &\text{card}\, \mathbf{S}_* \le l < 2 \cdot (\text{card}\, \mathbf{S}_*), \qquad \text{card}\, \mathbf{S}_* = 3 \\ &3 \le 3 < 2 \cdot 3 \end{aligned} \tag{98}$$

For the generation of the signal which halts the whole combined machine we can use (or add) the rule

$$S_{41\,42\,43\cdots} \to \varepsilon\, S_{\boldsymbol{HALT}},\ S_{\boldsymbol{HALT}} \notin \mathbf{S}_* \tag{99}$$

Here again it is valid that for the finite number of tracing steps and, for each is lasting the finitely long time, both halt states of our double-Turing Machine, which is the Turing Machine, too, occur *in a finite time*.

p ♯♯♯♯	q ♯♯♯♯	**Instruction**	**Configuration**	C ♯♯♯♯	**Config. Type**	g ♯♯♯♯	**Gerneral Config. Type** G	m ♯♯♯♯
1	**1**	$s_0\ I\ s_1\ I\ R$	$\varepsilon B[\ s_0\ IIIBIIII]B\varepsilon$	**1**	$\varepsilon\boldsymbol{B}[\ s_0\ \boldsymbol{IBI}]\boldsymbol{B}\varepsilon$	**1**	$\overrightarrow{\boldsymbol{B}}\ s_0\ \overrightarrow{\boldsymbol{IB}}$	**1**
2	2	$s_1\ I\ s_1\ I\ R$	$\varepsilon B[\ s_1\ IIBIIII]B\varepsilon$	2	$\varepsilon\boldsymbol{B}[\ s_1\ \boldsymbol{IBI}]\boldsymbol{B}\varepsilon$	2	$\overrightarrow{\boldsymbol{B}}\ s_1\ \overrightarrow{\boldsymbol{IB}}$	2
3	2	$s_1\ I\ s_1\ I\ R$	$\varepsilon B[I\ s_1\ IBIIII]B\varepsilon$	3	$\varepsilon\boldsymbol{B}[\boldsymbol{I}\ s_1\ \boldsymbol{IBI}]\boldsymbol{B}\varepsilon$	3	$\overrightarrow{\boldsymbol{BI}}\ s_1\ \overrightarrow{\boldsymbol{IB}}$	3
4	3	$s_1\ B\ s_2\ B\ R$	$\varepsilon B[II\ s_1\ BIII]B\varepsilon$	4	$\varepsilon\boldsymbol{B}[\boldsymbol{I}\ s_1\ \boldsymbol{BI}]\boldsymbol{B}\varepsilon$	4	$\overrightarrow{\boldsymbol{BI}}\ s_1\ \overrightarrow{\boldsymbol{BIB}}$	4
5	4	$s_2\ I\ s_2\ I\ R$	$\varepsilon B[IIB\ s_2\ III]B\varepsilon$	5	$\varepsilon\boldsymbol{B}[\boldsymbol{IB}\ s_2\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	5	$\overrightarrow{\boldsymbol{BIB}}\ s_2\ \overrightarrow{\boldsymbol{IB}}$	5
6	4	$s_2\ I\ s_2\ I\ R$	$\varepsilon B[IIBI\ s_2\ II]B\varepsilon$	6	$\varepsilon\boldsymbol{B}[\boldsymbol{IBI}\ s_2\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	6	$\overrightarrow{\boldsymbol{BI}}\ s_2\ \overrightarrow{\boldsymbol{IB}}$	6
7	4	$s_2\ I\ s_2\ I\ R$	$\varepsilon B[IIBII\ s_2\ I]B\varepsilon$	7	$\varepsilon\boldsymbol{B}[\boldsymbol{IBI}\ s_2\ \boldsymbol{I}]\boldsymbol{B}\varepsilon$	6	.....	6
8	5	$s_2\ B\ s_3\ B\ L$	$\varepsilon B[IIBIII\ s_2\ ]B\varepsilon$	8	$\varepsilon\boldsymbol{B}[\boldsymbol{IBIBI}\ s_2\ ]\boldsymbol{B}\varepsilon$	7	$\overrightarrow{\boldsymbol{BI}}\ s_2\ \overrightarrow{\boldsymbol{B}}$	7
9	6	$s_3\ I\ s_4\ B\ L$	$\varepsilon B[IIBII\ s_3\ IB]\varepsilon$	9	$\varepsilon\boldsymbol{B}[\boldsymbol{IBI}\ s_3\ \boldsymbol{IB}]\varepsilon$	8	$\overrightarrow{\boldsymbol{BI}}\ s_3\ \overrightarrow{\boldsymbol{IB}}$	8
10	6	$s_3\ I\ s_4\ B\ L$	$\varepsilon B[IIBIII\ s_3\ BB]\varepsilon$	10	$\varepsilon\boldsymbol{B}[\boldsymbol{IBI}\ s_3\ \boldsymbol{B}]\varepsilon$	9	$\overrightarrow{\boldsymbol{BI}}\ s_3\ \overrightarrow{\boldsymbol{B}}$	9
11	8	$s_4\ I\ s_5\ I\ L$	$\varepsilon B[IIBII\ s_4\ IBB]\varepsilon$	11	$\varepsilon\boldsymbol{B}[\boldsymbol{IBI}\ s_4\ \boldsymbol{IB}]\varepsilon$	10	$\overrightarrow{\boldsymbol{BI}}\ s_4\ \overrightarrow{\boldsymbol{IB}}$	10
12	9	$s_5\ I\ s_5\ I\ L$	$\varepsilon B[IIBI\ s_5\ IIBB]\varepsilon$	12	$\varepsilon\boldsymbol{B}[\boldsymbol{IBI}\ s_5\ \boldsymbol{IB}]\varepsilon$	11	$\overrightarrow{\boldsymbol{BI}}\ s_5\ \overrightarrow{\boldsymbol{IB}}$	11
13	9	$s_5\ I\ s_5\ I\ L$	$\varepsilon B[IIB\ s_5\ IIIBB]\varepsilon$	13	$\varepsilon\boldsymbol{B}[\boldsymbol{IB}\ s_5\ \boldsymbol{IB}]\varepsilon$	12	$\overrightarrow{\boldsymbol{BIB}}\ s_5\ \overrightarrow{\boldsymbol{IB}}$	12
14	10	$s_5\ B\ s_6\ B\ L$	$\varepsilon B[II\ s_5\ BIIIBB]\varepsilon$	14	$\varepsilon\boldsymbol{B}[\boldsymbol{I}\ s_5\ \boldsymbol{BIB}]\varepsilon$	13	$\overrightarrow{\boldsymbol{BI}}\ s_5\ \overrightarrow{\boldsymbol{BIB}}$	13
15	11	$s_6\ I\ s_6\ I\ L$	$\varepsilon B[I\ s_6\ IBIIIBB]\varepsilon$	15	$\varepsilon\boldsymbol{B}[\boldsymbol{I}\ s_6\ \boldsymbol{IBIB}]\varepsilon$	14	$\overrightarrow{\boldsymbol{BI}}\ s_6\ \overrightarrow{\boldsymbol{IB}}$	14
16	11	$s_6\ I\ s_6\ I\ L$	$\varepsilon B[\ s_6\ IIBIIIBB]\varepsilon$	16	$\varepsilon\boldsymbol{B}[\ s_6\ \boldsymbol{IBIB}]\varepsilon$	15	$\overrightarrow{\boldsymbol{B}}\ s_6\ \overrightarrow{\boldsymbol{IB}}$	15
17	12	$s_6\ B\ s_0\ B\ R$	$\varepsilon[\ s_6\ BIIBIIIBB]\varepsilon$	17	$\varepsilon[\ s_6\ \boldsymbol{BIBIB}]\varepsilon$	16	$\overrightarrow{\boldsymbol{B}}\ s_6\ \overrightarrow{\boldsymbol{BIB}}$	16
18	**1**	$s_0\ I\ s_1\ I\ R$	$\varepsilon[B\ s_0\ IIBIIIBB]\varepsilon$	**1**	$\varepsilon[\boldsymbol{B}\ s_0\ \boldsymbol{IBIB}]\varepsilon$	**1**	$\overrightarrow{\boldsymbol{B}}\ s_0\ \overrightarrow{\boldsymbol{IB}}$	**1**
19	2	$s_1\ I\ s_1\ I\ R$	$\varepsilon[B\ s_1\ IBIIIBB]\varepsilon$	$2'$	$\varepsilon[\boldsymbol{B}\ s_1\ \boldsymbol{IBIB}]\varepsilon$	2	$\overrightarrow{\boldsymbol{B}}\ s_1\ \overrightarrow{\boldsymbol{IB}}$	2
20	3	$s_1\ B\ s_2\ B\ R$	$\varepsilon[BI\ s_1\ BIIIBB]\varepsilon$	$4'$	$\varepsilon[\boldsymbol{BI}\ s_1\ \boldsymbol{BIB}]\varepsilon$	3	$\overrightarrow{\boldsymbol{BI}}\ s_1\ \overrightarrow{\boldsymbol{BIB}}$	3
21	4	$s_2\ I\ s_2\ I\ R$	$\varepsilon[BIB\ s_2\ IIIBB]\varepsilon$	$5'$	$\varepsilon[\boldsymbol{BIB}\ s_2\ \boldsymbol{IB}]\varepsilon$	5	$\overrightarrow{\boldsymbol{BIB}}\ s_2\ \overrightarrow{\boldsymbol{IB}}$	5
22	4	$s_2\ I\ s_2\ I\ R$	$\varepsilon[BIBI\ s_2\ IIBB]\varepsilon$	$6'$	$\varepsilon[\boldsymbol{BIBI}\ s_2\ \boldsymbol{IB}]\varepsilon$	5	$\overrightarrow{\boldsymbol{BI}}\ s_2\ \overrightarrow{\boldsymbol{IB}}$	6
23	4	$s_2\ I\ s_2\ I\ R$	$\varepsilon[BIBII\ s_2\ IBB]\varepsilon$	$7'$	$\varepsilon[\boldsymbol{BIBI}\ s_2\ \boldsymbol{IB}]\varepsilon$	6	.....	6
24	5	$s_2\ B\ s_3\ B\ L$	$\varepsilon[BIBIII\ s_2\ BB]\varepsilon$	$8'$	$\varepsilon[\boldsymbol{BIBI}\ s_2\ \boldsymbol{B}]\varepsilon$	7	$\overrightarrow{\boldsymbol{BI}}\ s_2\ \overrightarrow{\boldsymbol{B}}$	7
25	6	$s_3\ I\ s_4\ B\ L$	$\varepsilon[BIBII\ s_3\ IBB]\varepsilon$	$9'$	$\varepsilon[\boldsymbol{BIBI}\ s_3\ \boldsymbol{IB}]\varepsilon$	8	$\overrightarrow{\boldsymbol{BI}}\ s_3\ \overrightarrow{\boldsymbol{IB}}$	8
26	8	$s_4\ I\ s_5\ I\ L$	$\varepsilon[BIBI\ s_4\ IBBB]\varepsilon$	$11'$	$\varepsilon[\boldsymbol{BIBI}\ s_4\ \boldsymbol{IB}]\varepsilon$	10	$\overrightarrow{\boldsymbol{BI}}\ s_4\ \overrightarrow{\boldsymbol{IB}}$	10
27	9	$s_5\ I\ s_5\ I\ L$	$\varepsilon[BIB\ s_5\ IIBBB]\varepsilon$	$13'$	$\varepsilon[\boldsymbol{BIB}\ s_5\ \boldsymbol{IB}]\varepsilon$	12	$\overrightarrow{\boldsymbol{BIB}}\ s_5\ \overrightarrow{\boldsymbol{IB}}$	12
28	10	$s_5\ B\ s_6\ B\ L$	$\varepsilon[BI\ s_5\ BIIBBB]\varepsilon$	$15'$	$\varepsilon[\boldsymbol{BI}\ s_5\ \boldsymbol{BIB}]\varepsilon$	13	$\overrightarrow{\boldsymbol{BI}}\ s_5\ \overrightarrow{\boldsymbol{BIB}}$	13
29	11	$s_6\ I\ s_6\ I\ L$	$\varepsilon[B\ s_6\ IBIIBBB]\varepsilon$	$16'$	$\varepsilon[\boldsymbol{B}\ s_6\ \boldsymbol{IBIB}]\varepsilon$	15	$\overrightarrow{\boldsymbol{B}}\ s_6\ \overrightarrow{\boldsymbol{IB}}$	15
30	12	$s_6\ B\ s_0\ B\ R$	$\varepsilon[\ s_6\ BIBIIBBB]\varepsilon$	$17'$	$\varepsilon[\ s_6\ \boldsymbol{BIBIB}]\varepsilon$	16	$\overrightarrow{\boldsymbol{B}}\ s_6\ \overrightarrow{\boldsymbol{BIB}}$	16
31	**1**	$s_0\ I\ s_1\ I\ R$	$\varepsilon[B\ s_0\ IBIIBBB]\varepsilon$	**1**	$\varepsilon[\boldsymbol{B}\ s_0\ \boldsymbol{IBIB}]\varepsilon$	**1**	$\overrightarrow{\boldsymbol{B}}\ s_0\ \overrightarrow{\boldsymbol{IB}}$	**1**

p ♯♯♯♯	q ♯♯♯♯	**Instruction**	**Configuration**	C ♯♯♯♯	**Config. Type**	g ♯♯♯♯	**Gerneral Config. Type** G	m ♯♯♯♯
31	**1**	$s_0\ I\ s_1\ I\ R$	$\varepsilon[B\ s_0\ IBIIBBB]\varepsilon$	**1**	$\varepsilon[\boldsymbol{B}\ s_0\ \boldsymbol{IBIB}]\varepsilon$	**1**	$\vec{\boldsymbol{B}}\ s_0\ \vec{\boldsymbol{IB}}$	**1**
32	3	$s_1\ B\ s_2\ B\ R$	$\varepsilon[BB\ s_1\ BIIBBB]\varepsilon$	17	$\varepsilon[\boldsymbol{B}\ s_1\ \boldsymbol{BIB}]\varepsilon$	17	$\vec{\boldsymbol{B}}\ s_1\ \vec{\boldsymbol{BIB}}$	17
33	4	$s_2\ I\ s_2\ I\ R$	$\varepsilon[BBB\ s_2\ IIBBB]\varepsilon$	18	$\varepsilon[\boldsymbol{B}\ s_2\ \boldsymbol{IB}]\varepsilon$	18	$\vec{\boldsymbol{B}}\ s_2\ \vec{\boldsymbol{IB}}$	18
34	4	$s_2\ I\ s_2\ I\ R$	$\varepsilon[BBBI\ s_2\ IBBB]\varepsilon$	19	$\varepsilon[\boldsymbol{BI}\ s_2\ \boldsymbol{IB}]\varepsilon$	19	$\vec{\boldsymbol{BI}}\ s_2\ \vec{\boldsymbol{IB}}$	6
35	5	$s_2\ B\ s_3\ B\ L$	$\varepsilon[BBBII\ s_2\ BBB]\varepsilon$	20	$\varepsilon[\boldsymbol{BI}\ s_2\ \boldsymbol{IB}]\varepsilon$	19	.....	6
36	6	$s_3\ I\ s_4\ B\ L$	$\varepsilon[BBBI\ s_3\ IBBB]\varepsilon$	21	$\varepsilon[\boldsymbol{BI}\ s_3\ \boldsymbol{IB}]\varepsilon$	20	$\vec{\boldsymbol{BI}}\ s_3\ \vec{\boldsymbol{IB}}$	8
37	8	$s_4\ I\ s_5\ I\ L$	$\varepsilon[BBB\ s_4\ IBBBB]\varepsilon$	22	$\varepsilon[\boldsymbol{B}\ s_4\ \boldsymbol{IB}]\varepsilon$	21	$\vec{\boldsymbol{B}}\ s_4\ \vec{\boldsymbol{IB}}$	19
38	10	$s_5\ B\ s_6\ B\ L$	$\varepsilon[BB\ s_5\ BIBBB]\varepsilon$	23	$\varepsilon[\boldsymbol{B}\ s_5\ \boldsymbol{BIB}]\varepsilon$	22	$\vec{\boldsymbol{B}}\ s_5\ \vec{\boldsymbol{IB}}$	20
39	12	$s_6\ B\ s_0\ B\ R$	$\varepsilon[B\ s_6\ BBIBBB]\varepsilon$	24	$\varepsilon[\boldsymbol{B}\ s_6\ \boldsymbol{BIB}]\varepsilon$	23	$\vec{\boldsymbol{B}}\ s_6\ \vec{\boldsymbol{BIB}}$	21
40	12	$s_6\ B\ s_0\ B\ R$	$\varepsilon[BB\ s_6\ BIBBB]\varepsilon$	25	$\varepsilon[\boldsymbol{B}\ s_6\ \boldsymbol{BIB}]\varepsilon$	23	$\vec{\boldsymbol{B}}\ s_6\ \vec{\boldsymbol{BIB}}$	21
41	**13**	$s_0\ B\ s_0\ B\ L$	$\varepsilon[B\ s_0\ BBIBBBB]\varepsilon$	**26**	$\varepsilon[\boldsymbol{B}\ s_0\ \boldsymbol{BIB}]\varepsilon$	**24**	$\vec{\boldsymbol{B}}\ s_0\ \vec{\boldsymbol{BIB}}$	**22**
42	**13**	$s_0\ B\ s_0\ B\ L$	$\varepsilon B[\ s_0\ BBBIBBB]\varepsilon$	**26′**	$\varepsilon\boldsymbol{B}[\ s_0\ \boldsymbol{BIB}]\varepsilon$	**24**	.....	**22**
43	**13**	$s_0\ B\ s_0\ B\ L$	$\varepsilon B[\ s_0\ BBBBIBBB]\varepsilon$	**26″**	$\varepsilon\boldsymbol{B}[\ s_0\ \boldsymbol{BIB}]\varepsilon$	**24**	.....	**22**
44	**13**	$s_0\ B\ s_0\ B\ L$	$\varepsilon B[\ s_0\ BBBBBIBBB]\varepsilon$	**26‴**	$\varepsilon\boldsymbol{B}[\ s_0\ \boldsymbol{BIB}]\varepsilon$	**24**	.....	**22**
45	**13**	.....	....	**26⁗**	....	**24**	.....	**22**
..	..	.....	....	..	....	..	.....	..
..	**13**	.....	....	..	....	**24**	.....	**22**
....	**13**	$s_0\ B\ s_0\ B\ L$	$\varepsilon B[\ s_0\ B...B...BIBBB]\varepsilon$	**26′⋯′**	$\varepsilon\boldsymbol{B}[\ s_0\ \boldsymbol{BIB}]\varepsilon$	**24**	$\vec{\boldsymbol{B}}\ s_0\ \vec{\boldsymbol{BIB}}$	**22**
..	**13**	.....	....	..	....	**24**	.....	**22**
..	..	.....	....	..	....	..	.....	..

Table 3. Tracing and staging for Example III

[The distance Δp of p for the general configuration numbered **22** is constant ($\Delta p = 0$).]

Example IV

$$
\begin{aligned}
&\vec{\eta} = \left(\eta_1,\ \eta_2,\ \eta_3,\ \eta_4,\ \eta_5,\ \eta_6,\ \eta_7\ \ \eta_8,\ \eta_9,\ \eta_{10}\ \ \eta_{11},\ \eta_{12},\ \eta_{13}\right)\\
&\eta_1 = \left(s_0,\ I,\ s_1,\ B,\ R\right)\\
&\eta_2 = \left(s_1,\ I,\ s_1,\ I,\ R\right)\\
&\eta_3 = \left(s_1,\ B,\ s_2,\ B,\ R\right)\\
&\eta_4 = \left(s_2,\ I,\ s_2,\ I,\ R\right)\\
&\eta_5 = \left(s_2,\ B,\ s_3,\ B,\ L\right)
\end{aligned}
\tag{100}
$$

$$\eta_6 = (s_3,\ I,\ s_4,\ B,\ L)$$
$$\eta_7 = (s_4,\ B,\ s_H,\ B,\ R)$$
$$\eta_8 = (s_4,\ I,\ s_5,\ I,\ L)$$
$$\eta_9 = (s_5,\ I,\ s_5,\ I,\ L)$$
$$\eta_{10} = (s_5,\ B,\ s_6,\ B,\ L)$$
$$\eta_{11} = (s_6,\ I,\ s_6,\ I,\ L)$$
$$\eta_{12} = (s_6,\ B,\ s_0,\ B,\ R)$$
$$\eta_{13} = (s_0,\ B,\ s_0,\ B,\ L)$$

This program is figuring the difference 4-3 and the input is $\vec{\xi} = IIIBIIII$.

From the following table follows that during our staging the whole double-machine halts itself. In the numbers m of the general configuration types the following word $\mathbf{G}$ is generated,

$$\begin{aligned} &\mathbf{G} = [\mathbf{G}_1\mathbf{G}_2],\ l = 2 \\ &\mathbf{G}_1 = [1\,2\,3\,4\,5\,6\,7\,8\,9\,10\,11\,12] \\ &\mathbf{G}_2 = [1\,2\,3\,5\,13\,14\,15\,16] \end{aligned} \tag{101}$$

We can write the regular grammar of the (regular) language of the general configuration types being generated by the process driven by $\vec{\eta}$,

$$\begin{aligned} S_0 &\to \mathbf{G}_1\, S_{1\,2\,3\,4\,5\,6\,7\,8\,9\,10\,11\,12\,13} \\ S_{1\,2\,3\,4\,5\,6\,7\,8\,9\,10\,11\,12\,13} &\to \mathbf{G}_2\, S_{14\,15\,16\,17\,18\,19\,20\,21} \end{aligned} \tag{102}$$

This grammar has the set $\mathbf{S}_*$ of the nonterminal symbols,

$$\mathbf{S}_* = \left\{S_0\ \ S_{1\,2\,3\,4\,5\,6\,7\,8\,9\,10\,11\,12\,13}\ \ S_{14\,15\,16\,17\,18\,19\,20\,21} \equiv S_{\boldsymbol{HALT}}\right\},\ \text{card}\,\mathbf{S}_* = 3 \tag{103}$$

The whole *staging is ended by the natural end* of the process in the *observed machine*.

[All computing and tracing is stopped by the *end of the observed process*.]

All the previous examples have shown that after the finite number of steps, each is interpreted in a finite time and thus we need the finite time only, the signal halting the whole double machine is generated in the *all cases*. The reason for generating the *halting signal* and as such, the recognition the finite or infinite cycle in the observed machine is quite visible from all our tracing tables. Our double machine is the Turing Machine, too.

p ♯♯♯♯	q ♯♯♯♯	**Instruction**	**Configuration**	C ♯♯♯♯	**Config. Type**	g ♯♯♯♯	**Gerneral Config. Type** G	m ♯♯♯♯
1	**1**	$s_0\ I\ s_1\ B\ R$	$\varepsilon[\ s_0\ IIIBII]B\varepsilon$	**1**	$\varepsilon[\ s_0\ \boldsymbol{IBI}]\varepsilon$	**1**	$\overrightarrow{\boldsymbol{B}}\ s_0\ \overrightarrow{\boldsymbol{IB}}$	**1**
2	2	$s_1\ I\ s_1\ I\ R$	$\varepsilon[B\ s_1\ IIBII]B\varepsilon$	2	$\varepsilon[\boldsymbol{B}\ s_1\ \boldsymbol{IBI}]\varepsilon$	2	$\overrightarrow{\boldsymbol{B}}\ s_1\ \overrightarrow{\boldsymbol{IB}}$	2
3	2	$s_1\ I\ s_1\ I\ R$	$\varepsilon[BI\ s_1\ IBII]B\varepsilon$	3	$\varepsilon[\boldsymbol{BI}\ s_2\ \boldsymbol{IBI}]\varepsilon$	3	$\overrightarrow{\boldsymbol{BI}}\ s_2\ \overrightarrow{\boldsymbol{IB}}$	3
4	3	$s_1\ B\ s_2\ B\ R$	$\varepsilon[BII\ s_1\ BII]B\varepsilon$	4	$\varepsilon[\boldsymbol{BI}\ s_2\ \boldsymbol{BI}]\varepsilon$	4	$\overrightarrow{\boldsymbol{BI}}\ s_2\ \overrightarrow{\boldsymbol{BIB}}$	4
5	4	$s_2\ I\ s_2\ I\ R$	$\varepsilon[BIIB\ s_2\ II]B\varepsilon$	5	$\varepsilon[\boldsymbol{BIB}\ s_2\ \boldsymbol{II}]\varepsilon$	5	$\overrightarrow{\boldsymbol{BIB}}\ s_2\ \overrightarrow{\boldsymbol{IB}}$	5
6	4	$s_2\ I\ s_2\ I\ R$	$\varepsilon[BIIBI\ s_2\ I]B\varepsilon$	6	$\varepsilon[\boldsymbol{BIBI}\ s_3\ \boldsymbol{I}]\varepsilon$	6	$\overrightarrow{\boldsymbol{BI}}\ s_3\ \overrightarrow{\boldsymbol{IB}}$	6
7	5	$s_2\ B\ s_3\ B\ L$	$\varepsilon[BIIBII\ s_2\ B]\varepsilon$	7	$\varepsilon[\boldsymbol{BIBI}\ s_3\ \boldsymbol{B}]\varepsilon$	7	$\overrightarrow{\boldsymbol{BI}}\ s_3\ \overrightarrow{\boldsymbol{B}}$	7
8	6	$s_3\ I\ s_4\ B\ L$	$\varepsilon[BIIBI\ s_3\ IB]\varepsilon$	8	$\varepsilon[\boldsymbol{BI}\ s_3\ \boldsymbol{IB}]\varepsilon$	8	$\overrightarrow{\boldsymbol{BI}}\ s_3\ \overrightarrow{\boldsymbol{IB}}$	8
9	8	$s_4\ I\ s_5\ I\ L$	$\varepsilon[BIIB\ s_4\ IBB]\varepsilon$	9	$\varepsilon[\boldsymbol{BIB}\ s_4\ \boldsymbol{IB}]\varepsilon$	9	$\overrightarrow{\boldsymbol{BIB}}\ s_4\ \overrightarrow{\boldsymbol{IB}}$	9
10	10	$s_5\ B\ s_6\ B\ L$	$\varepsilon[BII\ s_5\ BIBB]\varepsilon$	10	$\varepsilon[\boldsymbol{BI}\ s_5\ \boldsymbol{BIB}]\varepsilon$	10	$\overrightarrow{\boldsymbol{BI}}\ s_5\ \overrightarrow{\boldsymbol{BIB}}$	10
11	11	$s_6\ I\ s_6\ I\ L$	$\varepsilon[BI\ s_6\ IBIBB]\varepsilon$	11	$\varepsilon[\boldsymbol{BI}\ s_6\ \boldsymbol{IBIB}]\varepsilon$	11	$\overrightarrow{\boldsymbol{BI}}\ s_6\ \overrightarrow{\boldsymbol{IB}}$	11
12	11	$s_6\ I\ s_6\ I\ L$	$\varepsilon[B\ s_6\ IIBIBB]\varepsilon$	12	$\varepsilon[\boldsymbol{B}\ s_6\ \boldsymbol{BBIB}]\varepsilon$	12	$\overrightarrow{\boldsymbol{B}}\ s_6\ \overrightarrow{\boldsymbol{BIB}}$	12
13	12	$s_6\ B\ s_0\ B\ R$	$\varepsilon[B\ s_6\ BIIBIBB]\varepsilon$	13	$\varepsilon[\boldsymbol{B}\ s_6\ \boldsymbol{BIBIB}]\varepsilon$	13	$\overrightarrow{\boldsymbol{B}}\ s_6\ \overrightarrow{\boldsymbol{BIB}}$	12
14	**1**	$s_0\ I\ s_1\ B\ R$	$\varepsilon[BB\ s_0\ IIBIB]\varepsilon$	**1′**	$\varepsilon[\boldsymbol{B}\ s_0\ \boldsymbol{IBIB}]\varepsilon$	**1**	$\overrightarrow{\boldsymbol{B}}\ s_0\ \overrightarrow{\boldsymbol{IB}}$	**1**
15	2	$s_1\ I\ s_1\ I\ R$	$\varepsilon[BBB\ s_1\ IBIBB]\varepsilon$	2′	$\varepsilon[\boldsymbol{B}\ s_1\ \boldsymbol{IBIB}]\varepsilon$	2′	$\overrightarrow{\boldsymbol{B}}\ s_1\ \overrightarrow{\boldsymbol{IB}}$	2
16	3	$s_1\ B\ s_2\ B\ R$	$\varepsilon[BBBI\ s_1\ BIBB]\varepsilon$	4′	$\varepsilon[\boldsymbol{BI}\ s_1\ \boldsymbol{BIB}]\varepsilon$	14	$\overrightarrow{\boldsymbol{BI}}\ s_1\ \overrightarrow{\boldsymbol{BIB}}$	3
17	4	$s_2\ I\ s_2\ I\ R$	$\varepsilon[BBBIB\ s_2\ IBB]\varepsilon$	5′	$\varepsilon[\boldsymbol{BIB}\ s_2\ \boldsymbol{IB}]\varepsilon$	5′	$\overrightarrow{\boldsymbol{BIB}}\ s_2\ \overrightarrow{\boldsymbol{IB}}$	5
18	5	$s_2\ B\ s_3\ B\ L$	$\varepsilon[BBBIBI\ s_2\ BB]\varepsilon$	7′	$\varepsilon[\boldsymbol{BIBI}\ s_2\ \boldsymbol{B}]\varepsilon$	15	$\overrightarrow{\boldsymbol{BI}}\ s_2\ \overrightarrow{\boldsymbol{B}}$	13
19	6	$s_3\ I\ s_4\ B\ L$	$\varepsilon[BBBIB\ s_3\ IBB]\varepsilon$	8′	$\varepsilon[\boldsymbol{BIB}\ s_3\ \boldsymbol{IB}]\varepsilon$	16	$\overrightarrow{\boldsymbol{BIB}}\ s_3\ \overrightarrow{\boldsymbol{IB}}$	14
20	7	$s_4\ B\ s_H\ B\ L$	$\varepsilon[BBBI\ s_4\ BBBB]\varepsilon$	14	$\varepsilon[\boldsymbol{BI}\ s_4\ \boldsymbol{B}]\varepsilon$	17	$\overrightarrow{\boldsymbol{BI}}\ s_4\ \overrightarrow{\boldsymbol{B}}$	15
21 xx	x xx	$(s_{HALT},\ I) \notin D_{\overrightarrow{\eta}}$ $xxxxx$	$\varepsilon[BBB\ s_{HALT}\ IBBBB]\varepsilon$ $xxxxx$	**15** xx	$\varepsilon[\boldsymbol{B}\ s_{HALT}\ \boldsymbol{IB}]\varepsilon$ $xxxxx$	**18** xx	$\overrightarrow{\boldsymbol{B}}\ s_{HALT}\ \overrightarrow{\boldsymbol{IB}}$ $xxxxx$	**16** xx

Table 4. Tracing and staging for Example IV

7. Conclusion

The *unsolvable decision problems* are of two types. The first type of the problem is solvable but not with the objects and decision-counting methods we have at our disposal. The example is the unsolvability of the binomical equations in the real axis. But with the Complex Numbers Theory they are solvable describing the physical reality. The help is that the *imaginary axis* (the new dimension) has been introduced. The another example is the Great Fermat Theorem and its solution (Andrew Wiles 1993 and Ann. Math. 1995).

The second type of the unsolvable or *undecidable* problems are those which are given mistakenly by *having an Auto-Reference embedded*. They are the *paradoxes*, which invokes the infinite cycles when they are 'solved' and just for this they are reducible to the Halting Problem; their

solution without a certain 'step-aside' requires the Perpetuum Mobile functionality. This doesn't mean that a counting under their description is not performed physically but is not resultative in a finite time.

(Nevertheless we can want to have the *infinite cycle for technology purposes*, e.g. the *push-pull circuit*. Here the infinite cycle's functionality is *created intentionally* and, as such, the push-pull circuit is the example of the *recurring but not recursive counting*. Here the Auto-Reference is introduced intentionally, as a successive figuring method, creating an infinite sequence of wanted values.)

Generally, a sequence of states or figuring steps of solving a problem could be *divergent* or *constant* or *convergent*. The divergent and constant cases are felt as the *real example of the infinite cycle in the very sense of this term*. We can say, rather jokingly, that the convergent counting halts, even if it was in the infinity (e.g. the Newton method) and that the divergent counting doesn't halt even in the infinity, including now the constant sequence too - the model is the information transfer in an interrupted information transfer channel. When in the recurring counting the number of figuring steps is not given explicitly, then, the results from the successive steps must be compared. When it is set badly or is not set at all, the infinite cycle occurs and, by the algorithm's definition requiring resultativness, such a counting is errorneous. The flagrant example of the badly set task is the way in which the Gibbs Paradox arises - here it is the Auto-Reference embedded by not respecting the difference between what is measured (observed), and what is measuring (observing). We used it in the extreme case (with the complete mixing these two levels) as the physical model of all noetic paradoxes.

The aim of this paper was to detect the infinite cycle from its own characteristics. Our vision is that the *counting itself is of the physical character* and, as such, is *subjected to physical laws*, especially, to the II Principle of Thermodynamics. The infinite cycle is viewed as a certain type of an equilibrium state.[17] To await the finite-time end of such states is a paradoxical and, as such, unachievable wish. Nevertheless, all these cycles are representable by the Carnot Cycle (it is the infinite cycle conceptually) used as the thermodynamic model of a cyclic information transfer [6, 8, 10]. We see here the *growth* of thermodynamic entropy within the *whole isolated system* in which the cycle, or the information transfer, is running and we see the *constant* or *decreasing* thermodynamic entropy within its *working medium*, or within the transfer channel in the information-thermodynamic representation. From this point of view the aim to recognize any infinite cycle, to decide the Halting Problem, is solvable. The information-thermodynamic considerations were expressed in terms of the Automaton Theory The general configuration types of the observed Turing Machine were generated and the Pumping Lemma was used. The author believes that he has shown that problems given paradoxically, errorneously as for being resultant, have the Auto-Reference embedded both *in* the sense of the *objective* of the problem or *in* the sense of *the solving* the problem - the *Auto-Reference can be in the solving method while the very objective of the problem can be solvable*. The author's wish is that

17 The interesting is that the stability of an equilibrium state and of an atomic structure are similar. Without the natural radioactivity the end of atoms seems to 'be in the infinity', too.

the following claim could be considered as the theorem for recognizing, deciding of any infinite cycle:

Due to the fact that any infinite cycle starts at a finite time and, for the Control Unit of any Turing Machine is an finite-state automaton, **and due to the fact that the Pumping Lemma it is valid for the regular infinite and thus periodical language of the general configuration types of the observed Turing Machine, the Halting Problem is decidable**.

8. Appendix

We consider the basic types of chains of the terminal symbols on the input-output *TM* 's tape,

$\mathbf{I} \triangleq \vec{I}\vec{I} = I,\ \vec{I} = II,\ \vec{I} = III\ldots$

$\mathbf{B} \triangleq \vec{B}\vec{B} = B,\ \vec{B} = BB,\ \vec{B} = BBB, \underline{\mathbf{B} \triangleq \varepsilon}$

Further types are

$\mathbf{IB} \triangleq \vec{I}\vec{B};\ \ \vec{I}\vec{B} = IB,\ \vec{I}\vec{B} = IIB,\ \vec{I}\vec{B} = III\ldots B$
$\vec{I}\vec{B} = IBB,\ \vec{I}\vec{B} = IBBB\ldots$
$\vec{I}\vec{B} = III\ldots BBB\ldots$

$\mathbf{BI} \triangleq \vec{B}\vec{I};\ \ \vec{B}\vec{I} = BI,\ \vec{B}\vec{I} = BBI,\ \vec{B}\vec{I} = BBB\ldots I$
$\vec{B}\vec{I} = BII,\ \vec{B}\vec{I} = BIII\ldots$
$\vec{B}\vec{I} = BBB\ldots III\ldots$

$\mathbf{IBI} \triangleq \vec{I}\vec{B}\vec{I} = III\ldots BBB\ldots III\ldots I$

$\mathbf{BIB} \triangleq \vec{B}\vec{I}\vec{B} = BBB\ldots III\ldots BBB\ldots B$

$\mathbf{IB}B \triangleq \mathbf{IB}$

$\mathbf{I}B\mathbf{B} \triangleq \mathbf{IB},$

$\mathbf{I}I\mathbf{B} \triangleq \mathbf{IB}$

$I\,\mathbf{IB} \triangleq \mathbf{IB}$

$\vec{\mathbf{IB}} \triangleq \mathbf{IB}, \vec{\mathbf{IB}} \triangleq \mathbf{IBIB}\ldots\mathbf{IBIB}$

$\vec{\mathbf{IB}}B \triangleq \mathbf{IB}B, \vec{\mathbf{IB}}B \triangleq \mathbf{IBIB}\ldots\mathbf{IB}B = \mathbf{IBIB}\ldots\mathbf{IBIB} = \vec{\mathbf{IB}}$

The following equivalences of chains of the terminal symbols are considerable,

$\vec{\mathbf{IBB}} \triangleq \mathbf{IBIB}\ldots\mathbf{IBIBB} = \mathbf{IBIB}\ldots\mathbf{IBIB} = \vec{\mathbf{IB}}$

$\vec{\mathbf{IBI}} \triangleq \mathbf{IBIIBI}\ldots\mathbf{IBIIBI} = \mathbf{IBIB}\ldots\mathbf{IBI}$

$\vec{\mathbf{IBII}} \triangleq \vec{\mathbf{IBI}}$

$\vec{\mathbf{IBIB}} \triangleq \vec{\mathbf{IB}}$

$$\overrightarrow{\overrightarrow{\mathbf{IBIB}}} \triangleq \overrightarrow{\mathbf{IBIB}}, \quad \overrightarrow{\overrightarrow{\mathbf{IBIB}}} \triangleq \overrightarrow{\mathbf{IBI}}\overrightarrow{\mathbf{BI}}\mathbf{...}\overrightarrow{\mathbf{BI}}\mathbf{BI}\overrightarrow{\mathbf{BI}}\mathbf{BI}\overrightarrow{\mathbf{BI}}\mathbf{BI}\overrightarrow{\mathbf{BI}}\mathbf{BI}\overrightarrow{\mathbf{BI}}\mathbf{BI}\overrightarrow{\mathbf{BI}}\mathbf{...}\overrightarrow{\mathbf{IBI}}\overrightarrow{\mathbf{BI}}\mathbf{B} = \overrightarrow{\mathbf{IB}}$$

$$\mathbf{B}\mathit{B}\mathbf{I} \triangleq \mathbf{BI}$$

$$\overrightarrow{\mathbf{BI}} \triangleq \mathbf{BI}$$

$$\overrightarrow{\mathbf{BI}} \triangleq \mathbf{BIBI...BIBI}$$

$$\overrightarrow{\mathbf{BI}}\mathit{I} \triangleq \mathbf{BI}\mathit{I}, \quad \overrightarrow{\mathbf{BI}}\mathit{I} \triangleq \mathbf{BIBI...BI}\mathit{I} = \mathbf{BIBI...BIBI} = \overrightarrow{\mathbf{BI}}$$

$$\overrightarrow{\mathbf{BIB}} \triangleq \mathbf{BIBI...BIBIB} = \overrightarrow{\mathbf{BIB}}$$

$$\overrightarrow{\mathbf{BIB}} \triangleq \mathbf{BIBBIB...BIBBIB} = \mathbf{BIBI...BIB}$$

$$\overrightarrow{\mathbf{BIBI}} \triangleq \overrightarrow{\mathbf{BI}}$$

$$\overrightarrow{\mathbf{BIBB}} \triangleq \overrightarrow{\mathbf{BIB}}$$

$$\overrightarrow{\overrightarrow{\mathbf{BIBI}}} \triangleq \overrightarrow{\mathbf{BIBIBI}}, \quad \overrightarrow{\overrightarrow{\mathbf{BIBI}}} \triangleq \overrightarrow{\mathbf{BIBI}}\overrightarrow{\mathbf{BI}}\mathbf{...}\overrightarrow{\mathbf{IBI}}\overrightarrow{\mathbf{BI}}\mathbf{BI}\overrightarrow{\mathbf{BI}}\mathbf{BI}\overrightarrow{\mathbf{BI}}\mathbf{...}\overrightarrow{\mathbf{BI}}\mathbf{BI}\overrightarrow{\mathbf{BI}}\mathbf{BI} = \overrightarrow{\mathbf{BI}}$$

Acknowledgements

This work was supported by the grant of Ministery of Education of the Czech Republic MSM 6046137307.

Author details

Bohdan Hejna

Address all correspondence to: hejnab@vscht.cz

University of Chemistry and Technology, Department of Mathematics, Prague, Czech Republic

References

[1] Ajzerman, M. A. a kol. *Logika automaty a algoritmy*; Academia: Praha, 1971.

[2] Hejna, B. Generování tabulek přechodů konečných automatů. Ing. Diplomová práce, Praha, FEL ČVUT, Praha, 1973.

[3] Hejna, B.; Vajda, I. Information transmission in stationary stochastic systems. *AIP Conf. Proc.* 1999, *465* (1), 405–418. DOI: 10.1063/1.58272.

[4] Hejna, B. Informační kapacita stacionárních fyzikálních systémů. Ph.D. Dissertation, ÚTIA AV ČR, Praha, FJFI ČVUT, Praha, 2000.

[5] Hejna, B. Tepelný cyklus a přenos informace. *Matematika na vysokých kolách: Determinismus a chaos*; JČMF, ČVUT: Praha, 2005; pp 83–87.

[6] Hejna, B. Thermodynamic Model of Noise Information Transfer. In *AIP Conference Proceedings*, Computing Anticipatory Systems: CASYS'07 – Eighth International Conference; Dubois, D., Ed.; American Institute of Physics: Melville, New York, 2008; pp 67–75. ISBN 978-0-7354-0579-0. ISSN 0094-243X.

[7] Hejna, B. Informační význam Gibbsova paradoxu. In *Matematika na vysokých kolách: Variačn principy*; JČMF: Praha, 2007; pp 25–31.

[8] Hejna, B. Gibbs Paradox as Property of Observation, Proof of II. Principle of Thermodynamics. In *AIP Conf. Proc.*, Computing Anticipatory Systems: CASYS'09: Ninth International Conference on Computing, Anticipatory Systems, 3–8 August 2009; Dubois, D., Ed.; American Institute of Physics: Melville, New York, 2010; pp 131–140. ISBN 978-0-7354-0858-6. ISSN 0094-243X.

[9] Hejna, B. *Informační termodynamika I.: Rovnovážná termodynamika přenosu informace*; VŠCHT Praha: Praha, 2010. ISBN 978-80-7080-747-7.

[10] Hejna, B. *Informační termodynamika II.: Fyzikální systémy přenosu informace*; VŠCHT Praha: Praha, 2011. ISBN 978-80-7080-774-3.

[11] Hejna, B. Information Thermodynamics, *Thermodynamics - Physical Chemistry of Aqueous Systems*, Juan Carlos Moreno-Piraján (Ed.), ISBN: 978-953-307-979-0, InTech, 2011. Available from: http://www.intechopen.com/articles/show/title/information-thermodynamics

[12] Hejna, B. Recognizing the Infinite Cycle: A Way of Looking at the Halting Problem. Lecture on CASYS'11 Conference, August 8-13, 2011, Liege, Belgium, *Proceedings of the Tenth International Conference CASYS'11 on Computing Anticipatory Systems, Lie'ge, Belgium, August 8-13, 2011, D. M. Dubois (Ed.), Publ. by CHAOS, 2012. ISSN 1373-5411.*

[13] Hejna, B. Information Capacity of Quantum Transfer Channels and Thermodynamic Analogies, *Thermodynamics - Fundamentals and its Application in Science*, Ricardo Morales-Rodriguez (Ed.), ISBN: 978-953-51-0779-8, InTech, 2012. Available from: http://www.intechopen.com/articles/show/title/information-capacity-of-quantum-transfer-channels-and-thermodynamic-analogies

[14] Hejna, B. *Informační Termodynamika III.: Automaty, termodynamika, přenos informace, výpočet a problém zastavení*; VŠCHT Praha, 2013, ISBN 978-80-7080-851-1.

[15] Hopcroft, J. E.; Ullman J. D. *Formálne jazyky a automaty*; Alfa: Bratislava, 1978.

[16] Horák, Z.; Krupka, F. *Technická fyzika*; SNTL/ALFA: Praha, 1976.

[17] Kalčík, J.; Sýkora, K. *Technická termomechanika*; Academia: Praha, 1973.

[18] Manna, Z. *Matematická teorie programů*; SNTL: Praha, 1981.

[19] Mareš, J. *Technická kybernetika*; ČVUT: Praha, 1992, 1997.

[20] Maršák, Z. *Termodynamika a statistická fyzika*; ČVUT: Praha, 1995.

[21] Melichar, Z. *Gramatiky a jazyky*; ČVUT: Praha, 1979, 1982.

[22] Moisil, Br. C. *Algebraická teorie automatů*; ČSAV: Praha, 1964.

[23] Prchal, J. *Signály a soustavy*; SNTL/ALFA: Praha, 1987.

[24] Kobrinskij N. E.; Trachtěnbrot B. A. *Úvod do teori*e konečných automatů; SNTL: Praha, 1967.

Chapter 7

Thermodynamics of Coral Diversity — Diversity Index of Coral Distributions in Amitori Bay, Iriomote Island, Japan and Intermediate Disturbance Hypothesis

Shinya Shimokawa, Tomokazu Murakami, Akiyuki Ukai, Hiroyoshi Kohno, Akira Mizutani and Kouta Nakase

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/61064

Abstract

The relationship between coral distributions and physical variables was investigated in Amitori Bay, Iriomote Island, Japan. Amitori Bay is located in the northeast region of Iriomote Island, Japan. Broad areas of coral have developed in the bay, and their life forms, coverages, sizes, and species vary depending on their locations. In addition, Amitori Bay has no access roads, and the bay perimeter is uninhabited. Thus, this small bay, with its variety of environments and lack of human impact, is considered to be one of the most suitable areas for studying the relationship between coral distribution and physical variables.

Field observations were conducted to obtain data on coral distributions, sea temperature, sea salinity, wind speed, and river flow rate. The observed data were then used in ocean and wave model numerical simulations and soil particle tracking analysis to obtain the spatial and temporal distributions of wave height and the numbers of soil particles. Our results showed that the life forms, sizes, and species of corals significantly varied depending on their locations in the bay, because the physical variables differed significantly among these locations.

From the results of the above observations and simulations, we calculated diversity index of coral distributions and its relation to physical variables. The diversity index, DI is defined as $DI=-\Sigma c_i \log_2 c_i$, where ci is the ratio of i-th group coverage to total coverage. DI is a quantitative measure for the degree in which a dataset includes different types and is related closely to entropy concept in Thermodynamics. The value of DI increases when both the number of types and the evenness increase. For a given number of types, the value of DI is maximized when all types are equally abundant. The results show that Averages of diversity index of the coral types at the mouth and inner parts of the bay are lower than average of the whole region, but average of diversity index at the intermedi-

ate part of the bay with the intermediate physical disturbances is higher than it. This seems to support the intermediate disturbance hypothesis which states species diversity in local area is maximized when environmental disturbances is neither too weak not too strong.

Keywords: Coral life form, Diversity index, Information entropy, Intermediate disturbance hypothesis

1. Introduction

"Corals" is the general term for *cnidaria* which inhabit the tropical and sub-tropical oceans and have an external skeleton [1, 2]. They are classified as either reef-building or non-reef-building corals. "Reef-building corals" is the general term for the species which inhabit the shallow oceans and have a relatively hard skeleton. They are the focus of this study. "Non-reef-building corals" is the general term for the species which inhabit the deep oceans and have a relatively soft skeleton. They include the so-called red, or precious, corals. Reef-building coral characteristics are deeply related to zooxanthellae existence. Zooxanthellae are a type of phytoplankton called dinoflagellate (mainly marine protozoa having two flagella). All reef-building corals live symbiotically with zooxanthellae. Instead of obtaining a secure habitat from reef-building corals, zooxanthellae contribute nine-tenths of their photosynthetic energy production to such corals. Reef-building corals can form their skeleton more rapidly and abundantly than non-reef-building corals by applying the photosynthetic energy production obtained from zooxanthellae. Reef-building coral skeletons are used as habitats by various organisms, and are the raw material for coral reefs. Reef-building corals also form mucus by applying the photosynthetic energy production not used for skeleton formation. The mucus plays a role in protecting the body surface of reef-building corals. Various organisms living in coral reefs treat them as a food. In other words, reef-building corals supply a large proportion of the habitats and foods for organisms in coral reefs, and thereby support abundant ecological systems in oligotrophic tropical oceans. Moreover, humans benefit in various ways from reef-building corals and coral reefs. Corallite, solidified from the skeletons of reef-building corals, is used for traditional daily necessaries and stone walls, and abundant sea foods obtained around coral reefs support our dietary requirements. Beautiful underwater views of coral reefs are treasured as tourist attractions around coastal areas of the tropics and sub-tropics. The function of coral reefs as a natural sea wall is important from disaster prevention viewpoint. In this study, "corals" is referred as "reef-building corals."

The distribution of corals is diverse and is affected by various factors such as human activities (including red soil erosion due to land development and sea temperature rise due to global warming), their predation by *Acanthaster* (crown-of-thorns starfish) and *Drupella fragum* (a species of sea snail), and environmental properties related to waves and soil grains. In these factors, environmental properties are the most important factors related to the long-term distribution of coral. Thus, investigating the relationship between coral distributions and environmental properties is useful in understanding coral and its diversity.

http://dx.doi.org/10.5772/61064

Some studies on the relationship have focused on wave height [3-7]. However, these studies did not consider spatial distribution. The relationship between wave forces and the spatial distribution of corals with a specific form in the Hawaiian Islands was investigated in [8]. However, their studies did not consider the effect of other environmental property such as soil grains. Recently, the relationships between coral habitat and some environmental properties were investigated in [9]. However, their studies focused only on the qualitative tendency.

This relationship includes short- and long-term physical processes along with direct responses to environmental properties. For example, in relationship to destruction by high waves, it is widely accepted that the structure of branching corals is more fragile than that of tabular corals. However, this destruction can also cause broader distribution due to the high regenerative power of corals. That is, high waves can reduce coral distribution in the short term but can broaden it in the long term. Conversely, extreme wave height, due to strong typhoons, can severely damage or destroy coral colonies [10, 4, 11, 12]. This seems to be related to the intermediate disturbance hypothesis (IDH) [13], which postulates that local species diversity is maximized when environmental disturbances are neither too weak (or rare) nor too strong (or frequent) and is often used to investigate the relationship between species diversity and environmental disturbances.

Environmental properties affect not only coral distribution but also coral life forms [1, 2]. Most corals make colonies consisting of many individuals called polyps. This type of coral is called a colonial coral, and all the corals in this study are classified as this type. Coral colonies have various forms such as tabular, branching, massive, foliose, and encrusting corals. They deform as per the environmental properties, and corals of the same species can show different forms. Reef-building coral classification is currently based on morphological features such as the shape of the individual polyps, their sequence manner on the colony, and the colony forms; however, classification is difficult because the surfaces of living corals have a molluscous covering [14]. In addition, although molecular phylogenetic analyses are needed for strict identification, there are cases when gene consistency within a species is lacking because gene exchanges among different species occur because of their unique mass-spawning trait (i.e., simultaneous release of eggs and sperm among different species) [15-17]. Currently, the right approach to investigate the relationship between coral distribution and environmental properties is considered to be one which regard habitat adaption as important, and one which focuses on the life form rather than the species.

Amitori Bay is located in the northeast region of Iriomote Island, Japan (Fig. 1), which is 2 km in width at its mouth, 4 km in length, and has a maximum depth of 70 m in the central mouth area, two rivers (the Ayanda and Udara) and the accompanied mangrove environments in the inner part of the bay. Broad areas of coral have been developed in the bay and their life forms, coverage, size, and species vary depending on their location. In addition, Amitori Bay has no access roads, and the bay perimeter is uninhabited. Thus, this small bay, with its variety of environments and lack of human impact, is considered to be one of the most suitable areas for studying the relationship between coral distribution and environmental properties.

On the contrary, observations of environmental properties in Amitori Bay are not many and are hard to conduct due to the access difficulty. Thus, we compensated for the shortage of

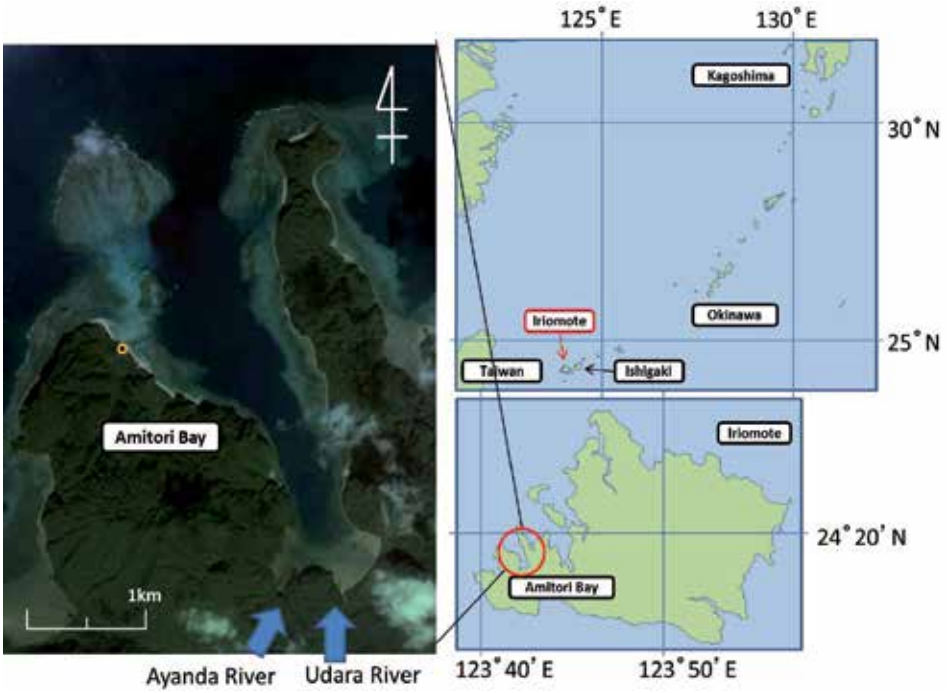

Figure 1. Locations of Amitori Bay and Iriomote Island. The top of the reef slope in Amitori Bay is indicated in Fig. 2 (after [18]).

observation by using numerical simulations. That is, when observational data were available, they are used to determine the parameters in the numerical simulations; on the other hand, when observational data were unavailable, methods without observational hypotheses in the numerical simulations are used.

This study [18] attempts to clarify these relationships in the corals of Amitori Bay, Iriomote Island, Japan, by observing the ocean, atmosphere, and rivers, which is performed through numerical simulation using ocean and wave models along with particle tracking analysis.

In Section 2, we state the investigation methods of the coral distribution and observation methods of the oceanic, atmospheric and river observations, and the numerical simulation methods to reproduce detailed oceanic and atmospheric states in Amitori Bay and to investigate the effects of waves and soil grains on coral distribution. In Section 3, we provide the results of our coral distribution investigation, our oceanic, atmospheric, and river observations, and the numerical simulations. In addition, we compare the coral distributions with our observational and numerical results, and discuss the relationship between coral distributions and environmental properties. In Section 4, we discuss the generality of the results obtained in Section 3 by conducting a statistical analysis. The analysis is related to diversity index, which is a quantitative measure for the degree to which a dataset includes different types and which is closely related to the information entropy concept. The results are also discussed from IDH perspective. We summarize and discuss our findings in Section 5.

http://dx.doi.org/10.5772/61064

2. Methods

2.1. Coral distribution investigation

Coral distributions were investigated at 44 locations in Amitori Bay, including 18 locations in 2009 and 26 locations in 2011 (Fig. 2). For the 2009 investigation, indicated by A–R in Fig. 2, three quadrats measuring 1 m on each side were placed at 1–4 points at various depths and life form, coverage, and coral size were recorded. For the 2011 investigation, indicated by 1–26 in Fig. 2, three quadrats measuring 2 m on each side were placed within a 3 m depth at the top of the reef slope and on the reef slope at a depth of 5–8 m, and life form, coverage, and coral size were recorded. The types of coral life forms treated in this study are shown in Fig. 3. States of the coral investigations in Amitori Bay and an example of the photographs of the quadrats are shown in Figs. 4 and 5, respectively. An electronic weighing instrument was used to determine the weight of trace pieces cut from the photographs of the quadrats and then, the coverage of each coral type was calculated. In this study, "coral individuals" and "coral size" mean the individuals and maximum diameters of a coral colony for a specific coral, respectively.

2.2. Oceanic and atmospheric observations

Fixed-point observations were conducted to obtain data on sea level, sea pressure, and horizontal current velocities in Amitori Bay from July 2008 to October 2009 using a WH-403 wave height/wave direction/current speed measuring instrument (I. O. Technic Co. Ltd.). Two measuring instruments were placed at each of Stations 1 and 2 at depths of 10 and 20 m, respectively, and measurements were conducted for 20-min periods at 1-h intervals. Also, moving shipboard observations were conducted to obtain sea temperature and salinity in the bay on October 18, 19, and 24; November 9 and 22; and December 14, 2011 using a RINKO conductivity, temperature, and depth profiler (JFE Advantech Co. Ltd.). The number of observation sites varied between 15 and 31 depending on the weather conditions. At each location, observations were conducted from the sea surface to the sea floor at 0.1-m intervals. At Station Z (Fig. 2), continuous measurements have been conducted for wind speed, wind direction, humidity, insolation, air temperature, sea surface temperature, and rainfall amount since 1976 [19]. During the study, the measurement interval was 10 min except for sea surface temperature measurement, which was 2 min.

2.3. River observation

To obtain the flow rates of two rivers, the Ayanda and Udara (Fig. 2) from July 21 to October 3, 2011, an AEM213-D electromagnetic current meter (JFE Advantech Co. Ltd.) were placed in an upstream stretch of each river. Water depths and cross-sectional area were also obtained at each location. The rain volumes were measured at Station Z. Then, a regression model equation was constructed to estimate the flow rate from the rainfall volume. This was performed because the measurement of flow rates throughout the year was difficult due to the need for ongoing instrument maintenance. The correlation coefficients between the calculated results and the observations were high over 0.8.

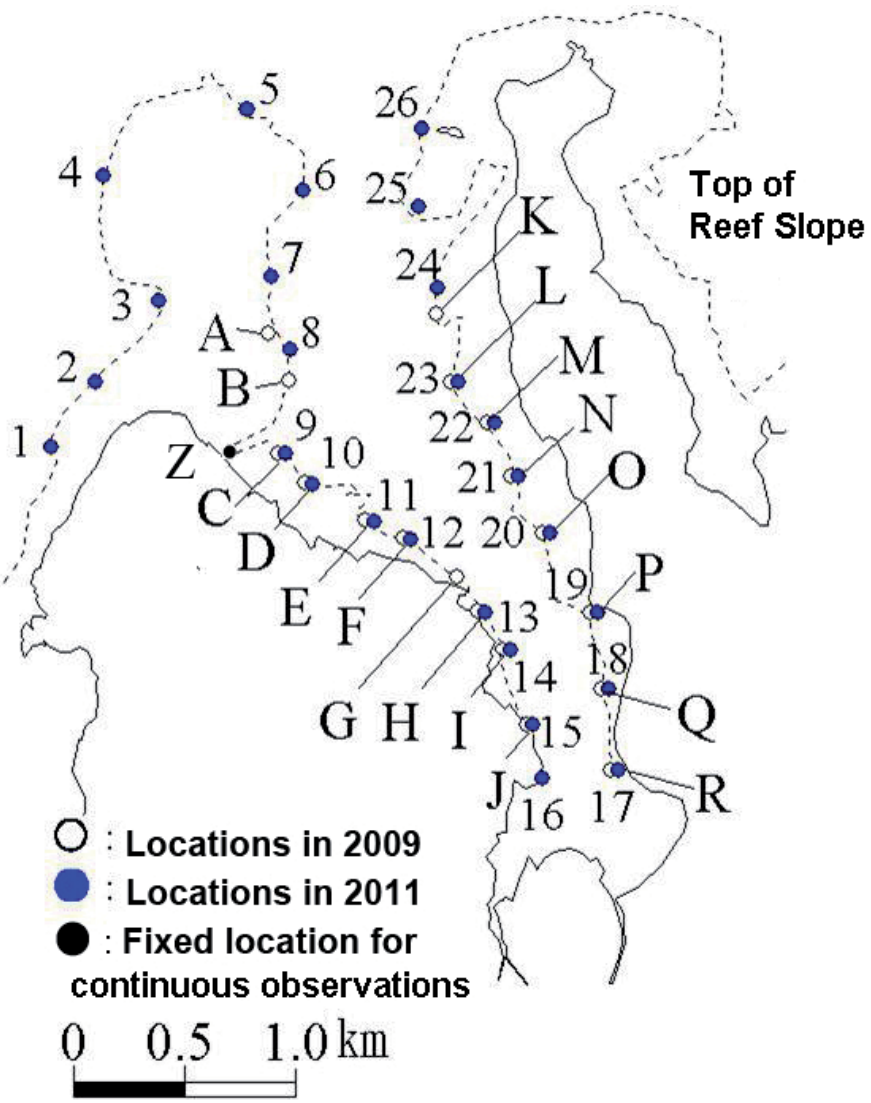

Figure 2. Locations for coral distribution investigations in Amitori Bay. Stations A–R and 1-26 show the locations in 2009 and 2011, respectively. Station Z shows the location for continuous measurements (position: 24°19′51.6″N, 123°41′21.6″E, and height: 4.2 m; after [18]).

2.4. Wave simulation

Numerical simulations were conducted to calculate wave heights and directions in Amitori Bay according to [7]. The calculation was conducted during a one-year period from November 1, 2008 to October 31, 2009 and in the following two regions: the large region including Yaeyama islands of 150 km × 100 km with 500-m grid spacing and the small region (i.e., Amitori Bay) of 4.5 km × 3 km with 20-m grid spacing. Offshore wave conditions from Japan Mete-

http://dx.doi.org/10.5772/61064

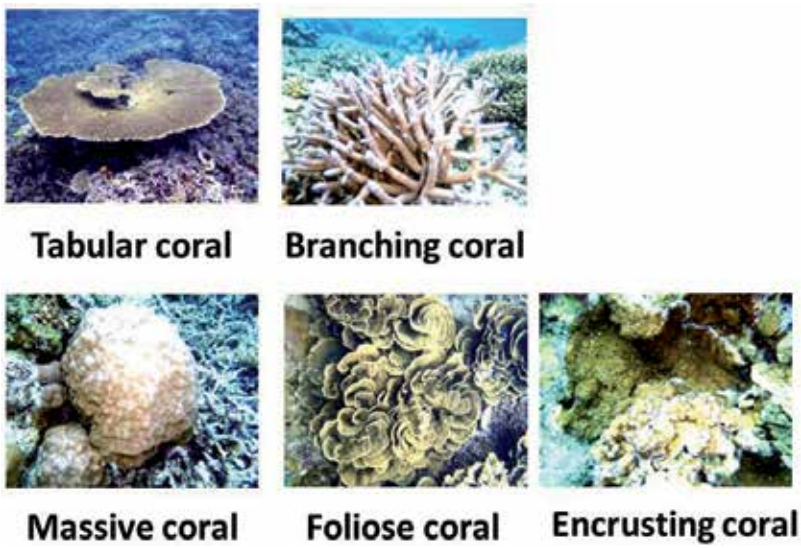

Figure 3. Types of coral life forms treated in this study (after [18]).

orological Agency Grid Point Value [20] such as wave heights, directions and periods were used as input data for the calculation. Three basic wave heights and directions were determined on the basis of the offshore wave conditions. Nine case calculations, as combinations of the basic wave heights and directions, were conducted using an energy balance equation and a wave breaking model developed by [21]. Then, a transformation table was created from the calculation results for every mesh in Amitori Bay, which was used to combine wave heights and directions in Amitori Bay with offshore wave conditions of wave height, direction and period. By using this table, time series of wave heights and directions in Amitori Bay during the target term were obtained from the corresponding offshore wave conditions.

2.5. Soil grain simulation

Lagrangian particle tracking analysis was conducted to obtain the transport properties of soil grains in Amitori Bay. For the purpose, numerical simulations were conducted to reproduce the flow fields in Amitori Bay from 1600 JST on October 18 to 0400 JST on October 25, 2011 using the coastal ocean current model (CCM). The details of the model are described in [22, 23]. Observational data (temperature, salinity, wind speed, wind direction, humidity, insolation, air temperature and rainfall amount), the flow rate from the regression model and astronomical tides calculated by the North Atlantic Oscillation oceanic tide model [24] were used as initial and boundary values for the numerical simulations. The reproductively of the observations in the model was confirmed with high precision [18]. The details of the simulation methods are described in [18]. The particle tracking analysis was conducted using the above flow fields calculated from the CCM and assuming the following conditions: (1) sedimentation of a soil grain due to the own weight was calculated using Rubey's [25] experimental equation, (2) there is no diffusion effect, (3) soil grains that reached the sea floor were not fixed but

Figure 4. Photographs of coral investigations in Amitori Bay.

continued to flow when the Shields number exceeded a critical value, and (4) calcareous sediment produced by the reefs themselves was not considered. The Shields number represents mobility of soil grains on sea or river bottoms, and the critical value was set to 0.05 [26].

http://dx.doi.org/10.5772/61064

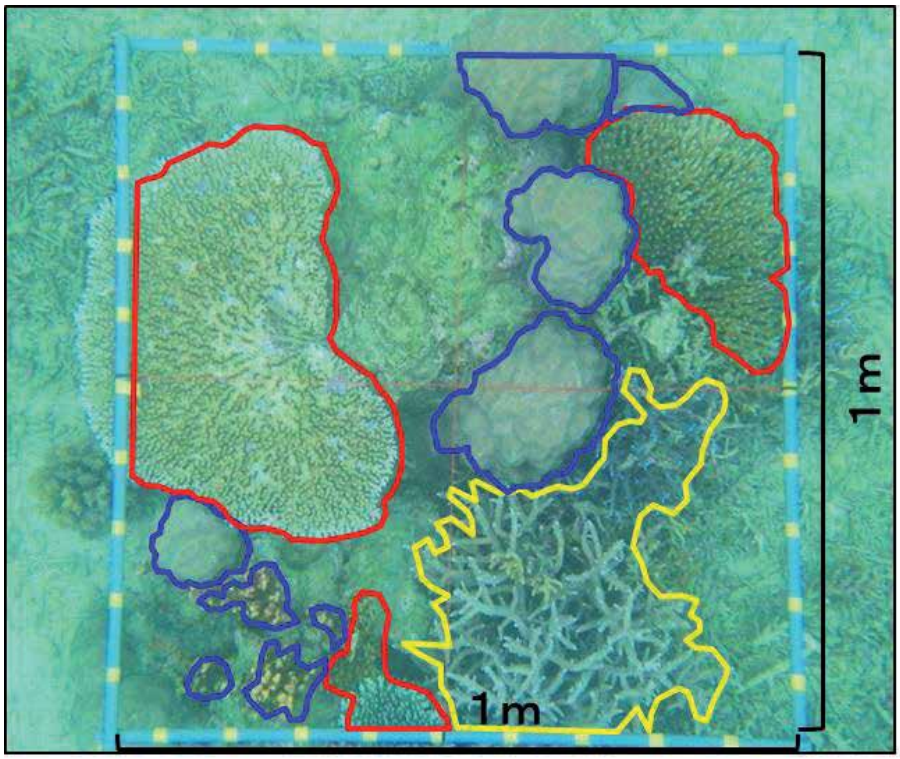

Figure 5. An example of photographs of quadrats. Red, yellow and blue lines show tabular, branching and massive corals, respectively.

Diameters of soil grains were set to 0.1, 1, 3, 5, 8, 10, 15, 20, and 30 μm because sediment trap observations at the mouths of the rivers showed that the diameters were distributed from 0.1 to 50 μm [18]. They were released every 15 s from the Ayanda and Udara rivers.

2.6. Focus of our study and representativeness for the normal state

In this sub-section, we highlight a few points relating to the focus of our study and the representativeness of the normal state. Further details are given by [18].

1. In considering the effect of river inflows on corals in the bay, we focused on soil grains rather than salinity because the effect of low salinity due to inflows from the rivers is smaller and shorter than that of soil grains. In considering the effect of oceanic mechanical force on corals in the bay, we focused on wave height rather than tidal current because tidal full and residual currents have a smaller effect on corals than does wave height.

2. The periods of oceanic simulation (October 18–25, 2011) and soil grain tracking analysis (June 2–22 and November 1–21, 2011) were considered to be in a normal state during these months at Iriomote Island. This was determined by comparison between observational data during the target periods and those for the most recent 33 years.

3. The normal wave state of the bay is calm compared with that of open oceans. Wave heights in the bay are only slightly affected by offshore wave direction because of the relationship between the overhanging coral reefs at the mouth of the bay, which limit the wave direction to the north, and the most prevalent wind direction. For the same reasons, the effect of waves due to typhoons on corals in the bay is small.

3. Relationship between coral distributions and environmental properties

3.1. Distribution of coral life forms

Fig. 6 shows the ratios of coverage of each coral type in Amitori Bay during the 2009 and 2011 investigations classified according to life form. Tabular and branching corals were dominant, covering a total area of more than 90%. The former were dominant at the top of the reef slopes, and the latter were dominant on the reef slopes. On the east side of the bay, the coverage of tabular corals tended to decrease from the mouth of the bay (Station 26) to the inner bay (Station 17) though the tendency was not clear on the west side of the bay. These results suggest that Amitori Bay, although small, has various types of corals.

3.2. Relationship between coral distributions and wave heights

Fig. 7a shows the spatial distribution of wave heights with a corresponding 95% probability of non-exceedance calculated by the wave simulation described in Section 2.4. The wave heights in the mouth of the bay (Stations 4 and 26) were considerably higher than those in the inner bay (Station 17) and on the east side of the bay (Station 20). Fig. 7b shows the relationship between the coverage of tabular or branching corals and wave height. The results showed that a larger coverage of tabular coral and smaller coverage of branching coral corresponded to larger wave heights. These results suggest that tabular coral is strengthened by high wave disturbances and that branching coral is easily broken by them, which is consistent with the results of previous studies [27, 28].

3.3. Relationship between coral distributions and the number of soil grains

Fig. 8 shows the number of soil grains on the sea bottom calculated by the particle tracking analysis described in Section 2.5 and the coverages of tabular and branching corals at Stations 1–26. A comparison between the number of soil grains on the sea bottom (Fig. 8a) and the coverage of the tabular corals (Fig. 8b) showed that smaller coverage of the coral related to a larger number of grains. However, the same comparison with branching corals (Fig. 8c) showed that the coverage of the coral did not relate to the number of grains. These results suggest that tabular coral is affected by soil grains, but branching coral is not because soil grains are easier to accumulate on tabular coral than on branching coral because of their shape.

http://dx.doi.org/10.5772/61064

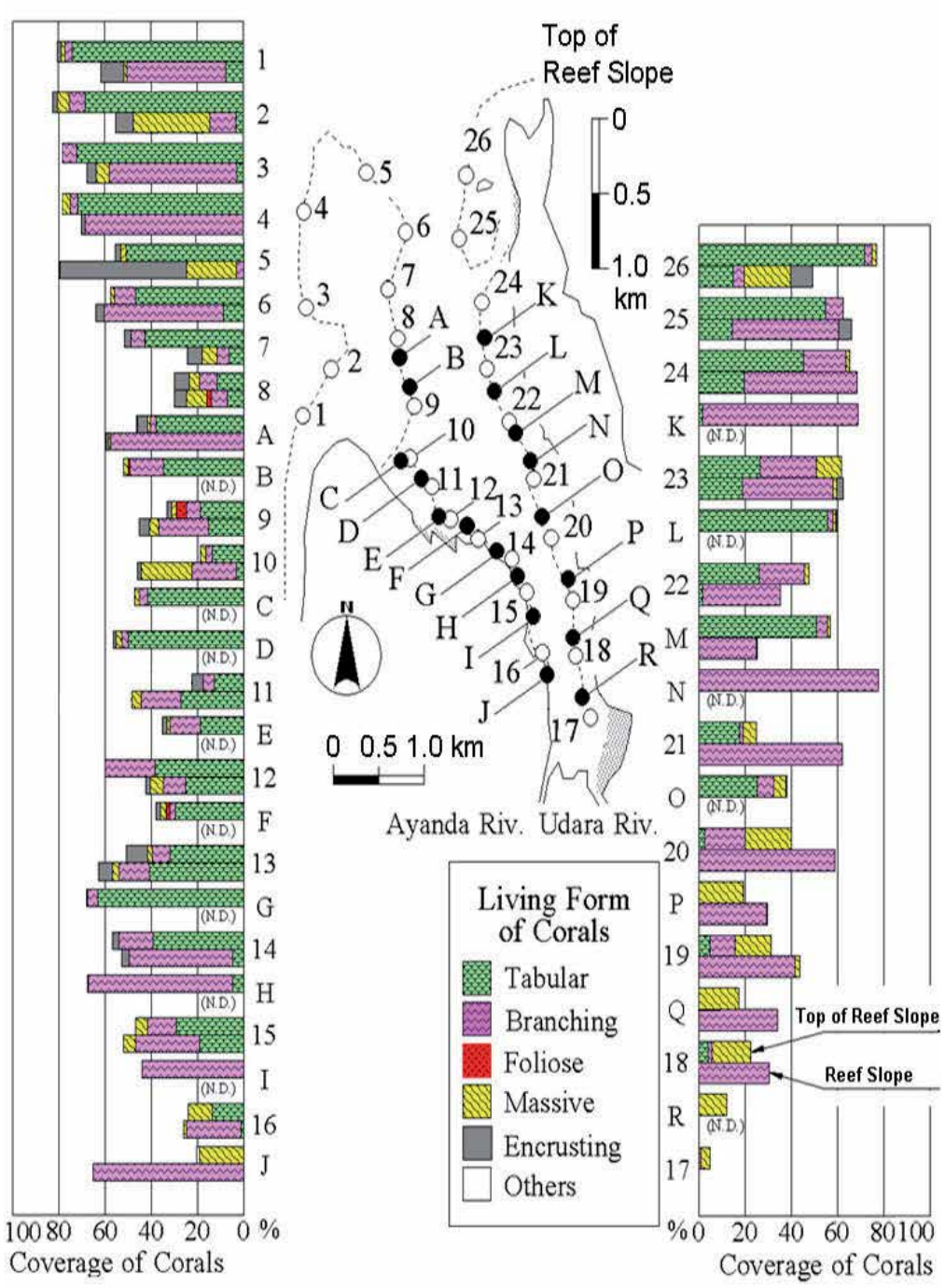

Figure 6. Ratios of coverage of each coral type in Amitori Bay during the 2009 and 2011 investigations classified according to life form, and the corresponding investigation locations. N.D. indicates no data on reef slope. Types of coral life forms treated in this study are shown in Fig. 3 (after [18]).

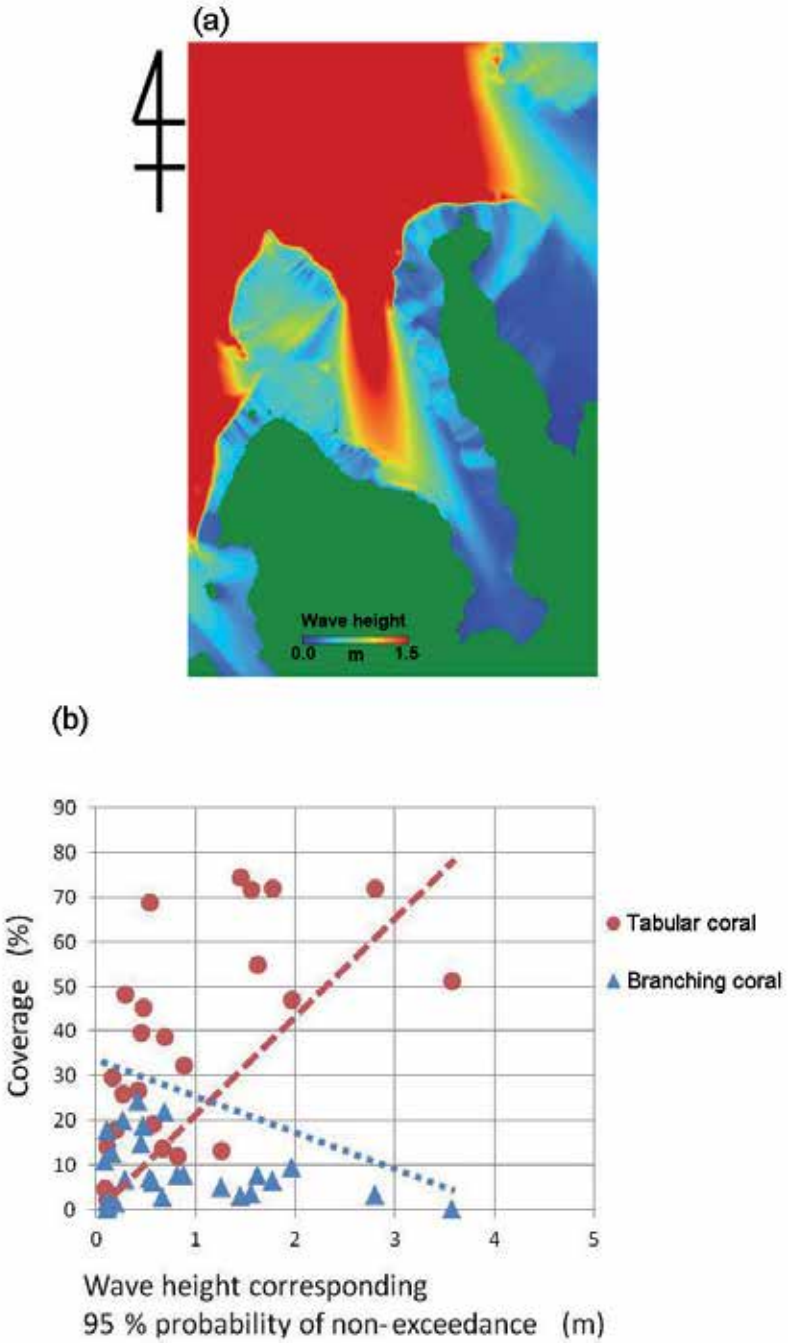

Figure 7. a) Spatial distribution of wave heights with a corresponding 95% probability of non-exceedance calculated by the wave simulation described in Section 2.4. (b) Relationship between the coverage of tabular or branching corals and wave height. The red- and blue-dotted lines show the lower boundary of coverage of the tabular corals and the upper boundary of coverage of the branching corals, respectively (after [18]).

http://dx.doi.org/10.5772/61064

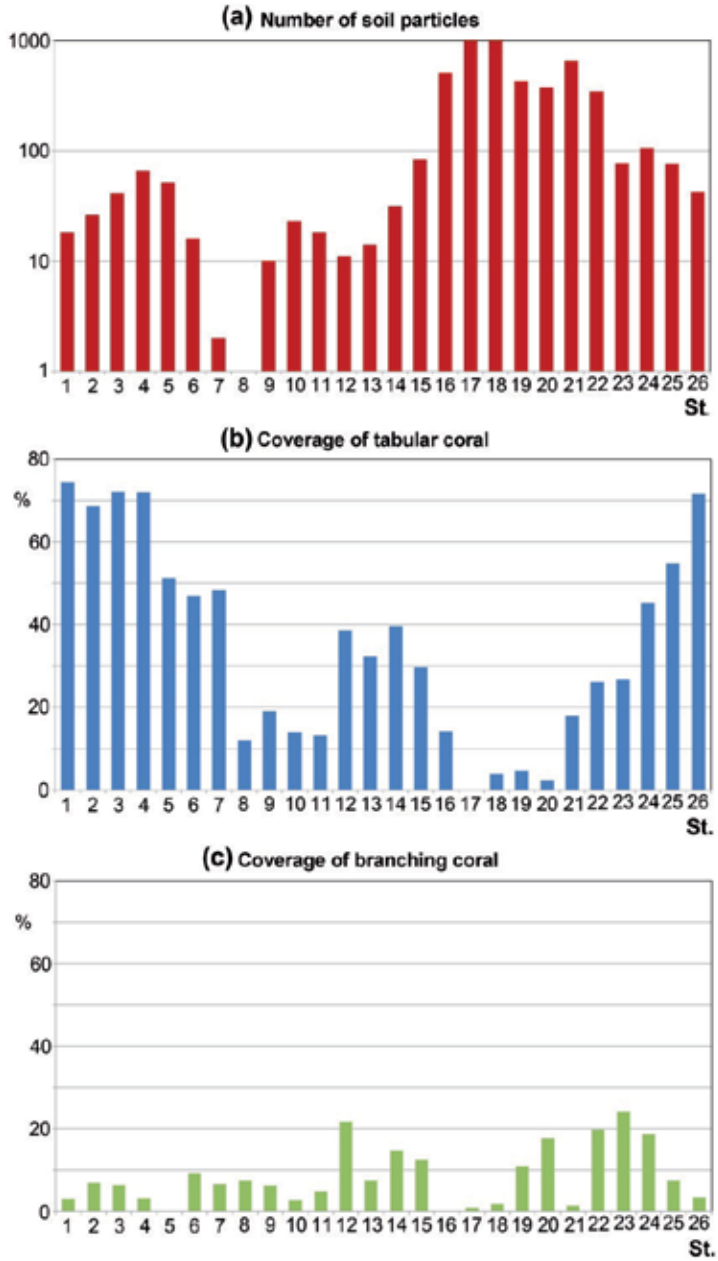

Figure 8. a) Number of soil grains that reached the sea floor by the end of the period (0400 JST on October 25, 2011) calculated by the particle tracking analysis described in section 2.5 and coverage of (b) tabular coral and (c) branching coral at Stations 1–26. The sizes of grains counted in (a) were 1, 3, and 5 μm in diameter. The number of soil grains at Stations 17 and 18 in (a) were over 1000 (after [18]).

4. Diversity index analysis

4.1. Diversity index and its physical meaning

In this section, the generality of the results is discussed by conducting a statistical analysis of the data obtained in the previous sections. The diversity index, H′ [29-32] is calculated and is defined as follows:

$$H' = -\sum c_i \log_2 c_i, \tag{1}$$

where c_i is the ratio of the i^{th} group population to the total population. H′ is a quantitative measure for the degree to which a dataset includes different types and is closely related to the information entropy concept. The value of H′ is larger than zero and has no upper limit. It increases when both the number of types and evenness increase.

For a given number of types, it is maximized when all types are equally abundant. For example, a case with 5 individuals of only 1 type, i.e., $H' = -(5/5) \log_2 (5/5) = 0.0$. In this case, the result is zero when a sample has been taken because the type of the sample is determined. Conversely, in a case with 1 individual of each of 5 types, $H' = -(1/5) \log_2 (1/5) \times 5 = 2.32$. In this case, the result is greater than zero when a sample has been taken because the type of the sample has multiple possibilities. In a case with a constant number of types, for example, cases with either 1 individual or 10 individuals of each of 5 types, i.e., ci = 1/5 = 10/50, and thereby H′ = 2.32. That is, in a case with a constant number of types and the same distribution of those types, H′ is unaffected by the number of individuals.

Conversely, in a case with a constant number of individuals, for example, a case with 2 individuals of each of 5 types, $H' = -(2/10) \log_2 (2/10) \times 5 = 2.32$, but in a case with 1 individual of each of 10 types, $H' = -(1/10) \log_2 (1/10) \times 10 = 3.32$. That is, in a case with a constant number of individuals, the greater the number of types, the greater is H′. Thus, H′ can be considered as an index that represents diversity and considers the distribution of individuals, not the individuals themselves.

Apart from this index, Simpson's λ index [33] represents diversity based on the number of types and individuals. In addition, Warwick and Clarke's $\Delta+$ and $\Delta-$ index [34] represents diversity by considering the taxonomic composition of the types present (i.e., taxonomic distinctness index). Refer to [31, 35] for the characteristics of these indices and the differences among them.

In our case, "group" refers to the coral type (Fig. 3); that is, the number of the group is five although "group" is originally defined as the species [30]. This is because we have focused on the coral type instead of the species as stated in the introduction. In addition, we used coverage instead of population because corals form colonies although ci is originally defined as the ratio of the i-th group population to the total population.

http://dx.doi.org/10.5772/61064

4.2. Distribution of the diversity index and the IDH

Fig. 9 shows the values of H′ in Amitori bay. The mouth, intermediate, and inner area of the bay are indicated in green, blue, and red, respectively. Averages of H′ at the mouth of the bay (0.60) and the inner area of the bay (0.62) were lower than the average value for the entire area (0.83). However, the average of H′ at the intermediate area of the bay (0.97) was higher than the average for the entire area. This distribution of H′ is due to the existence of large environmental disturbances, i.e., large wave height and large numbers of soil grains, in the mouth and inner area of the bay, respectively.

This seems to support the IDH demonstrated by [13], which postulates that local species diversity is maximized when environmental disturbances are neither too weak (or rare) nor too strong (or frequent) and is often used to investigate the relationship between species diversity and environmental disturbances. When the strength or frequency of disturbances to a biological community is low, competitive elimination of certain species by more dominant species in the community occurs, and the dominant species become the majority. When the strength or frequency of disturbances to a biological community is high, only specific species that are tolerant to the disturbances can survive, and these species become the majority. That is, when disturbances are either low or high, diversity decreases; thus, diversity is maximized when disturbances are intermediate. In addition, the IDH suggests that the environment is in a non-equilibrium state (i.e., due to the existence of disturbances and the subsequent recovery processes), which helps maintain diversity. In other words, if the environment is in a non-equilibrium state, frequent disturbances can provide a survival opportunity to those species which cannot survive in an equilibrium state; thereby, diversity is maintained.

For sites with good coral conditions, for example, Badul Island Waters, Ujung Kulon, Indonesia [36] and the Gulf of Aqaba and Ras Mohammed, Red Sea, Egypt [37], the averages of H′ were 2.183 and 0.84, respectively. H′, in this study (Fig. 9), seemed to be slightly lower than the reported values. However, a simple comparison of H′ among different area and/or studies is difficult because H′ is sensitive to the degree of sampling effort (e.g., the number of individuals and types) [31]. In this study, H′ was calculated from only five types because the object for calculating H′ was coral forms not coral species. We consider this explains the slightly lower value of H′ in our study.

4.3. Relationship between diversity index and environmental properties

Next, the relationships among H′, wave height (Fig. 10) and the number of soil grains (Fig. 11) are discussed. In general, the H′ value of organisms has a large dispersion. In addition, H′ is dependent on the number of individuals, and the dispersion of H′ decreases with an increase in the number of individuals [31]. Thus, the discussion of H′ should be performed using the maximum or average values. In the inner area of the bay with nearly constant (and low) wave height, no correlation between H′ and wave height was found. Conversely, in the intermediate area of the bay with changes in wave height, the maximum (or average) of H′ at a wave height increased with an increase in wave height up to 1 m. However, H′ decreased sharply when wave height was over 1 m (Fig. 10). Thus, it is considered that the diversity of coral types increases with an increase in wave disturbance; however, when the wave disturbances exceed

a limit, coral types which are vulnerable to such disturbances are limited, and the diversity decreases. Conversely, H′ peaked at average soil grain numbers of approximately 900 and 40,000 (Fig. 11), and the increased tendency of H′ to increase with disturbances (up to a limit) is unclear when compared to the relationship of H′ to significant wave height. H′ reached a maximum at an average soil grain number of approximately 900 at Stations 8 and 9 in the intermediate area of the bay. At Stations 8 and 9, the ratio of tabular coral was low; thus, the ratios of other corals, especially foliose coral, were high (Fig. 6). These characteristic were not seen at other locations but, at Stations 8 and 9, were considered to be the cause of high H′ values. Foliose coral may prefer conditions with low numbers of soil grains and intermediate wave heights, which are typical of the conditions at Stations 8 and 9. When Stations 8 and 9 are excluded from Fig. 11, the tendency of H′ to increase with an increase in disturbances (up to a limit in the average number of soil grains of approximately 40,000) becomes slightly clearer. The distribution of H′ in Fig. 9 is considered to be explained by these relationships of H′ to wave height and the number of soil grains.

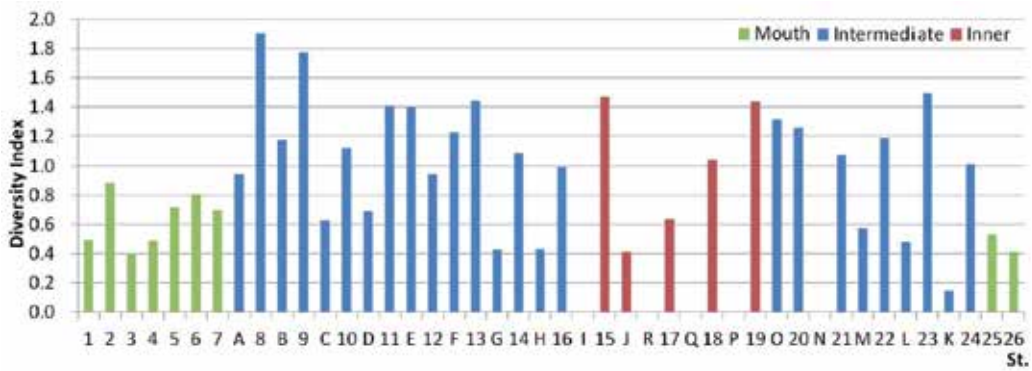

Figure 9. Diversity index at Stations 1–26 and A–R. The mouth, intermediate, and inner area of the bay are indicated in green, blue, and red, respectively (after [18]).

Note that high diversity does not indicate high coverage. Low diversity and low coverage indicate that the area is not a niche for organisms or biocoenosis distributed in the area. However, low diversity and high coverage indicate that the area is a niche only for organisms or biocoenosis distributed in the area. Fig. 12 shows the relationship between H′ and the coverage of tabular or branching corals. For both corals, a high H′ value (over 1.4), which is higher than the average of H′ (0.83), corresponds to coverage from 5% to 30%, and high coverage (over 50%) corresponds to a H′ value lower than 0.6, which is lower than the average of H′. An area with high diversity of corals is not considered to high coverage of a specific coral type (tabular or branching coral in this case) because the area is not a niche only for the species. An area with high coverage of a specific coral type is not considered to have high diversity of corals because the area is a niche only for the species.

http://dx.doi.org/10.5772/61064

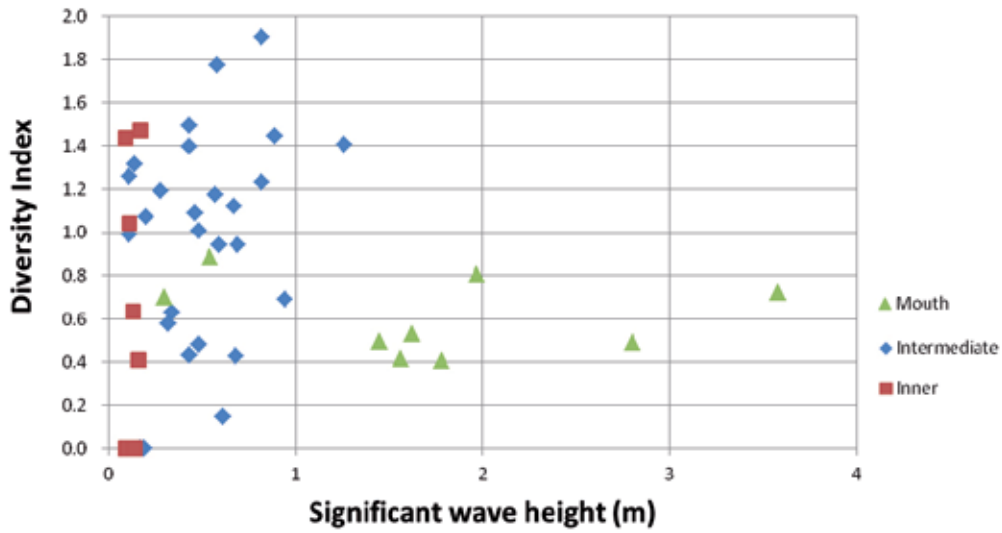

Figure 10. Relationship between diversity index and significant wave height at Stations 1–26 and A–R. The mouth, intermediate, and inner area of the bay are indicated in green, blue, and red, respectively (after [18]).

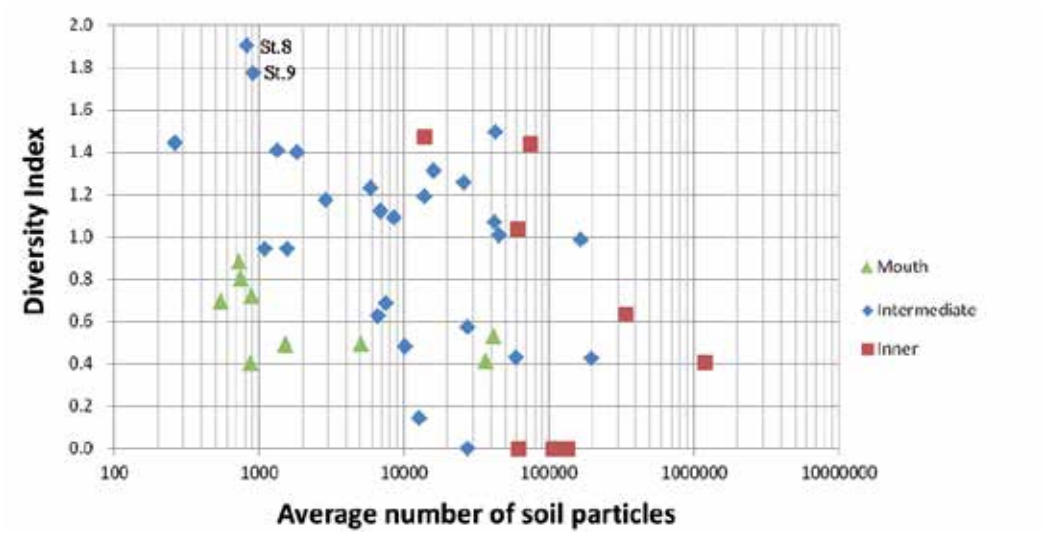

Figure 11. Relationship between diversity index and average number of soil grains reaching the sea floor per day at Stations 1–26 and A–R. The mouth, intermediate, and inner area of the bay are indicated in green, blue, and red, respectively. The numbers in the figure are the sum for all grain sizes (after [18]).

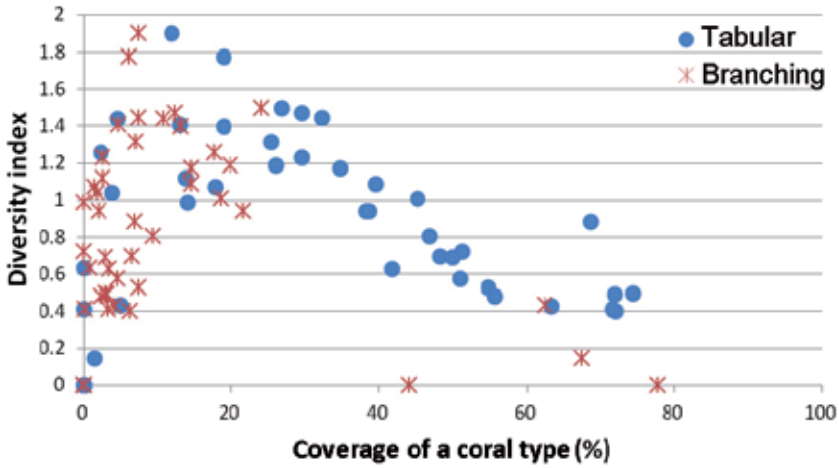

Figure 12. Relationship between diversity index and the coverage of tabular or branching corals at Stations 1–26 and A–R (after[18]).

5. Conclusion

In this study, field observations were conducted to obtain data on coral distributions, sea temperature, sea salinity, wind speed, and river flow rate. Ocean and wave model numerical simulations and soil grain tracking analysis were also conducted to obtain the spatial and temporal distributions of wave height and the number of soil grains. Using these observational and numerical data, the relationship between coral distributions and environmental properties in Amitori Bay, Iriomote Island, Japan, were investigated. Moreover, diversity indices (which are a quantitative measure for the degree to which a dataset includes different types and which are closely related to information entropy concept) of coral distributions were calculated and their relationships to environmental properties were investigated. Our results showed that the life forms, sizes, and species of corals significantly varied depending on their locations in the bay because the environmental properties differed significantly between these locations.

The main results of this study can be summarized as follows:

1. Coral distribution had a large correlation with wave height and number of soil grains. Larger coverage of tabular coral and smaller coverage of branching coral correlated with larger wave heights. Smaller coverage of tabular coral related to a larger number of soil grains, whereas branching coral coverage did not correlate with the number of soil grains.

2. Averages of diversity index of the coral types at the mouth and inner area of the bay with high environmental disturbances were lower than the average of the entire area. However, the average of the diversity index at the intermediate area of the bay with intermediate environmental disturbances was higher than it. This seems to support the IDH demon-

http://dx.doi.org/10.5772/61064

strated by [13], which postulates that local species diversity is maximized when environmental disturbances are neither too weak nor too strong.

Currently, many coral reefs around the world are endangered [1, 2]. Possible causes for this include red soil erosion (due to land development), sea temperature rise (due to global warming), and *Acanthaster* (crown-of-thorns starfish) and *Drupella fragum* (a species of sea snail) infestations. Red soil erosion causes a turbidity rise in sea water and soil adhesion to the coral surface, thereby hindering zooxanthellae photosynthesis and coral growth. Sea temperature rise causes zooxanthellae decolorization in the corals or discharge from the corals (i.e., coral breeching). Lengthening of coral breeching may cause coral death. Large outbreaks of *Acanthaster* and *Drupella fragum* prey on the molluscous portion of corals cause a direct decrease in coral numbers. The cause of large outbreaks of *Acanthaster* and *Drupella fragum* is regarded as human activities such as eutrophication (caused by inputs of domestic sewage). The corals in Amitori Bay have also been affected by various factors such as coral bleaching by high sea temperatures in 1997, 1998, and 2007; damage by *Acanthaster* infestations between 1980 and 1983; the recoveries from these incidents. To conserve corals in our oceans and regions such as Amitori Bay, the relationship between corals and their environmental properties must be clarified and continuously monitored. Further studies of this relationship in other coral regions are strongly advised.

Acknowledgements

We thank Prof. Kenshi Kimura of Tokai University for his helpful comments, Ms. Chiharu Yamaguchi of the National Research Institute for Earth Science and Disaster Prevention for helping in designing the figures and American Geophysical Union for permission to reuse the figures of [18] (all figures except for Figs.4 and 5 in this chapter). This study was supported by a corporative research project of the Okinawa Regional Research Center of Tokai University, Penta-Ocean Construction Co. Ltd, National Research Institute for Earth Science and Disaster Prevention, and the Japan Society for the Promotion of Science through Grant No. 25400465.

Author details

Shinya Shimokawa[1], Tomokazu Murakami[1], Akiyuki Ukai[2], Hiroyoshi Kohno[3], Akira Mizutani[3] and Kouta Nakase[2]

1 Storm, Flood, and Landslide Research Unit, National Research Institute for Earth Science and Disaster Prevention, Japan

2 Environment Business Division, Civil Engineering Headquarters, Penta-Ocean Construction. Co. Ltd., Japan

3 Okinawa Regional Research Center, Tokai University, Japan

References

[1] Motokawa, T. (2008). Story of Coral Reefs and Coral—A Strange Ecosystem of the Southern Ocean, Chuokoron-shinsya, Tokyo, pp. 273 (in Japanese).

[2] Sheppard, C.R.C.; Davy, S.K.; Pilling, G. M. (2009). The Biology of Coral Reefs, Oxford University Press, Oxford, pp. 339.

[3] Massel, S.R. & Done, T.J. (1993). Effects of Cyclone Waves on Massive Coral Assemblages on the Great Barrier Reef: Meteorology, Hydrodynamics and Demography, Coral Reefs, Vol. 12, No. 3-4, pp. 153-166.

[4] Madin, J.S. & Connolly, S.R. (2006). Ecological Consequences of Major Hydrodynamic Disturbances on Coral Reefs, Nature, Vol. 444, No. 7118, pp. 477-480.

[5] Toritani, M.; Sanuki, H.; Nakase, K.; Eriguchi, T. & Okayasu, A. (2006). Flow Environment Suitable to Coral Growth on Outer Block Structure in Naha bay, In Proceedings of the 9th Japanese Coral Reef Society (Ed. T. Nakamori, Japanese Coral Reef Society, Tokyo), p. 46 (in Japanese).

[6] Shibata, S.; Aota, T.; Mitsui, J.; Watanuki, A.; Kumagaya, W.; Nadaoka, K. & Taniguchi, H. (2007). Possible Effect of Waves and Currents on Coral Growth around Akijima Island, Midoriishi, Vol. 18, pp. 19-23 (in Japanese).

[7] Ukai, A., Kohno, H., Nakase, K., Shimaya, M., Zinno, M. & Kimura, K. (2010), Influence of Ocean Waves on Coral Habitat Environment of Amitori Bay, Annual Journal of Civil Engineering: Ocean, Vol. 26, pp. 363-368 (in Japanese with English abstract).

[8] Storlazzi, C.D.; Brown, E.K.; Field, M.E.; Rodgers, K.; Jokiel, P.L. (2005). A Model for Wave Control on Coral Breakage and Species Distribution in the Hawaiian Islands, Coral Reefs, Vol. 24, No. 1, pp. 43-55.

[9] Freeman, L.A.; Miller, A.J.; Norris, R. D. & Smith, J. E. (2012). Classification of Remote Pacific Coral Reefs by Physical Oceanographic Environment, Journal of Geophysical Research: Oceans, Vol. 117, No. C2, doi:10.1029/2011JC007099/full.

[10] Dollar, S. J. & Tribble, G.W. (1993). Recurrent Storm Disturbance and Recovery: A Long-term Study of Coral Communities in Hawaii, Coral Reefs, Vol. 12, No. 3-4, pp. 223-233.

[11] Hongo, C.; Kawamata, H. & Goto, K. (2012). Catastrophic Impact of Typhoon Waves on Coral Communities in the Ryukyu Islands under Global Warming, Journal of Geophysical Research: Biogeosciences, Vol. 117, No. G2, doi:10.1029/2011JG001902/full.

[12] White, K.N.; Ohara, T.; Fujii, T.; Kawamura, I.; Mizuyama, M.; Montenegro, J.; Shikiba, H.; Naruse, T.; McClelland, T.Y.; Denis, V. & Reimer, J. D. (2013). Typhoon Damage on a Shallow Mesophotic Reef in Okinawa, Japan, PeerJ, Vol. 1, e151, doi.org/10.7717/peerj.151.

[13] Connell, J. H. (1978). Diversity in Tropical Rain Forests and Coral Reefs, Science, Vol. 199, No. 4335, pp. 1302-1310.

[14] Fukami, H.; Tachikawa, H.; Suzuki, G.; Nagata, S. & Sugihara, K. (2010). Current Status and Problems with the Identification and Taxonomy of Zooxanthellate Scleractinian Corals in Japan, Journal of Japanese Coral Reef Society, Vol. 12, pp. 17-31 (in Japansese).

[15] Hatta, M.; Fukami, H.; Wang, W.; Omori, M.; Shimoike, K.; Hayashibara, T.; Ina, Y. & Sugiyama, T. (1999). Reproductive and Genetic Evidence for a Reticulate Evolutionary History of Mass-spawning Corals, Molecular Biology and Evolution, Vol. 16, No. 11, pp. 1607-1613.

[16] Fukami, H.; Budd, A.F.; Paulay, G.; Sole-Cava, A.; Chen, C.A.; Iwao, K. & Knowlton, N. (2004). Conventional Taxonomy Obscures Deep Divergence Between Pacific and Atlantic corals, Nature, Vol. 427, No. 6977, pp. 832-835.

[17] Fukami, H.; Chen, C.A.; Budd, A.F.; Collins, A.; Wallace, C.; Chuang, Y.-Y.; Chen, C.; Dai, C.-F.; Iwao, K.; Sheppard, C. & Knowlton, N. (2008). Mitochondrial and Nuclear Genes Suggest that Stony Corals are Monophyletic but most Families of Stony Corals are not (Order Scleractinia, Class Anthozoa, Phylum Cnidaria), PLoS One, Vol. 3, No. 9, e3222.

[18] Shimokawa, S.; Murakami, T.; Ukai, A.; Kohno, H.; Mizutani, A. & Nakase, K. (2014). Relationship Between Coral Distributions and Physical Variables in Amitori Bay, Iriomote Island, Japan, Journal of Geophysical Research: Oceans, Vol. 119, No. 12, pp. 8336-8356, doi: 10.1002/2014JC010307.

[19] Mizutani, A. & Sakihara, T. (2012). Atmospheric Observation in Okinawa Regional Research Center, Tokai Univerisity, Study Rev. Iriomote Is. 2011, pp. 84-95.

[20] Japan Meteorological Agency (JMA). (2015). http://gpvjma.ccs.hpcc.jp/~gpvjma/, Japan Meteorological Agency, Tokyo.

[21] Isobe, M. (1986). Calculation Method for Refraction, Diffraction and Breaking of Irregular Wave using Parabolic Equation, Annual Journal of Coastal Engineering, Vol. 33, pp. 134-138 (in Japanese).

[22] Murakami, T.; Yoshino, J.; Yasuda, T.; Iizuka, S. & Shimokawa, S. (2011). Atmosphere–Ocean–Wave Coupled Model Performing 4DDA with a Tropical Cyclone Bogussing Scheme to Calculate Storm Surges in an Inner Bay. Asian Journal of Environment and Disaster Management, Vol. 3, No. 2, pp. 217–228.

[23] Shimokawa, S.; Murakami, T.; Iizuka, S.; Yoshino, J. & Yasuda, T. (2014a). A New Typhoon Bogussing Scheme to Obtain the Possible Maximum Typhoon and its Application for Assessment of Impacts of the Possible Maximum Storm Surges in Ise and Tokyo Bays in Japan, Natural Hazards, Vol. 74, pp. 2037-2052, doi:10.1007/s11069-014-1277-2.

[24] Matsumoto, K.; Takanezawa, T. & Ooe, M. (2000). Ocean Tide Models Developed by Assimilating TOPEX/POSEIDON Altimeter Data into Hydrodynamical Model: A Global Model and a Regional Around Japan, Journal of Oceanography, Vol. 56, pp. 567-581.

[25] Rubey, W.W. (1933). Settling Velocities of Gravel, Sand and Silt Particles, American Journal of Science, Vol. 25, pp. 325-338.

[26] Kawasaki, K.; Murakami, T.; Toda, K. & Okubo, Y. (2008). Particle Tracking Analysis on Seawater Exchange and Soil Transport in Ise Bay Area at Tokai Heavy Rain, Annual Journal of Coastal Engineering, Vol. 55, pp. 986-990. (in Japanese with English abstract).

[27] Goreau, T.F. (1959). The Ecology of Jamaican Coral Reefs, 1. Species Composition and Zonation, Journal of Ecology, Vol. 40, pp. 67-90.

[28] Woodley, J.D.; Chornesky, E.A.; Clifford, P.A.; Jackson, J.B.C.; Kaufman, L.S.; Knowlton, N.; Lang, J.C.; Pearson, M.P.; Porter, J.W.; Rooney, M.C.; Rylaarsdam, K.W.; Tunnicliffe, V.J.; Wahle, C. M.; Wulff, J.L.; Curtis, A.S.G.; Dallmeyer, M. D.B.; Jupp, P.; Koehl, M.A.R.; Neigel, J.; Sides, E.M. (1981). Hurricane Allen's Impact on Jamaican Coral Reefs, Science, Vol. 214, pp. 749-755.

[29] Shannon, C.E. (1948). A Mathematical Theory of Communication. The Bell System Technical Journal, Vol. 27, pp. 379-423 and 623-656.

[30] MacArthur, R.H. & MacArthur, J.W. (1961). On Bird Species Diversity, Ecology, Vol. 42, No. 3, pp. 594-598.

[31] Clarke, K.R. & Warwick, R.M. (2001). Changes in Marine Communities: An Approach to Statistical Analysis and Interpretation, (2nd ed.) *Plymouth Marine Laboratory: PRIMER-E Ltd*, p. 176.

[32] McCune, B. & Grace, J.B. (2002). Analysis of Ecological Communities, Gleneden Beach, Oregon: MjM software design, Vol. 28, pp. 300.

[33] Simpson, E. H. (1949). Measurement of Diversity, Nature, Vol. 163, pp. 688-688.

[34] Warwick, R.M. & Clarke, K.R. (2001). Practical Measures of Marine Biodiversity Based on Relatedness of Species. Oceanography and Marine Biology, An Annual Review, Vol. 39, pp. 207-231.

[35] Ohgaki, S. (2008). Diversity and Similarity—New Taxonomic Index, *Argonauta*, Vol. 15, pp. 10-22 (in Japanese).

[36] Putri, L.S.E.; Hidayat, A.F. & Sukandar, P. (2012). Diversity of Coral Reefs in Badul Island Waters, Ujung Kulon, Indonesia, *Journal of Biological Sciences*, Vol. 1, pp. 59-64.

[37] Shokry, M. & Ammar, A. (2011). Coral Diversity Indices along the Gulf of Aqaba and Ras Mohammed, Red Sea, Egypt, *Biodiversitas*, Vol. 12, pp. 92-98.

Chapter 8

Thermodynamics of Abiotic Stress and Stress Tolerance of Cultivated Plants

Vesna Dragičević

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/60990

Abstract

Plants, as living systems, depend simultaneously on their internal status and their surroundings. Changes in plants' surroundings, generated by different environmental factors (abiotic stress), could perturb existing homeostasis, thus imposing stress. Abiotic stress includes heat, cold, freezing, flooding, drought (refers to water deficit), weak or strong light, oxygen deficiency or sufficiency, increased UV or other ionising rays, high salinity or acidity of the soil, deficiency or sufficiency of mineral elements, and presence of pollutants (xenobiotics). The effect of each abiotic factor depends on its severity, duration, developmental stage of the plant and its susceptibility to stress. During stress, requirements for energy increase (with increased intensity of respiration —domination of exergonic processes) as well as entropy. Variations in environmental factors could push the plant's metabolism out of homeostasis. In order to reestablish it, smaller or higher amounts of energy are required. The intention to increase the yield (grain or biomass production) of cultivated plants requires additional energy for successful completion of their life cycle, which makes them especially susceptible to stressful environments. From this point, the necessity to develop tolerant genotypes, which require less energy for maintaining homeostasis, arises.

Keywords: Agriculture, stress, tolerance, homeostasis

1. Introduction

Plants, as living systems, are self-organized systems depending simultaneously on their internal status and their surroundings. They transform solar energy and substances into

chemical energy during photosynthesis, along with newly synthesized substances of the plant organism itself (anabolism). Moreover, energy transformation, mainly during respiration (catabolism), considers some losses of free energy in the form of heat, increasing entropy. In order to maintain self-regulation and uphold growth and development, plant systems have to be open. This means that they are far removed from equilibrium. Homeostasis, representing the balance between entropy and enthalpy, with a steady inward flow of energy, is the most stable state that an open system can achieve.

However, changes in the surroundings could perturb existing homeostasis, thus imposing stress. This could be generated by different environmental factors (abiotic stress). Abiotic stress includes temperature increase or decrease (heat, cold and freezing stress), flooding, drought (as a combination of water deficiency and high temperatures), strong light, deficiency and/or sufficiency in CO_2 and O_2 (photosynthesis disturbance, respiration/photorespiration, hypoxia, and anoxia), increased UV or other ionising rays, increased levels of ozone, high salinity or acidity of the soil, deficiency and/or sufficiency of mineral elements, the presence of pollutants (heavy metals, persistent synthetic chemicals, some volatile organic and inorganic compounds, etc.), as well as agrochemicals (including pesticides), namely, xenobiotics. The susceptibility of plants to abiotic stress varies depending on the degree of stress, different accompanying stress factors, plant species, and their developmental stage.

Depending on intensity and/or duration, stress may be weak to moderate and short-term, with almost full recovery of plants in a much shorter period; or it could be intensive to severe and long-term, when recovery is highly limited and with plants permanently injured. Each stressful factor has a specific mode of action. Irrespectively of this, in every individual case, or under a combination of several factors, the same or a similar response could be generated in plants. For this reason, the hypothesis of the existence of a general adaptation syndrome was introduced [1]. Since plants are sessile organisms, mechanisms of tolerance (i.e., stress avoidance and stress adaptation) were developed. For cultivated plants, the most known mechanism of stress avoidance is the formation of seeds, bulbs, tubers, or other organs for vegetative rest, useful for surviving under extreme environmental conditions. When the mechanism of tolerance is activated, plants achieve distress metabolic pathways, which include reversible processes (i.e., recovery mechanisms), with lower energy consumption and less enthalpy variation. This means that tolerant plants have higher capacities to adjust or to adapt to the stress and thus, attain homeostasis. Adaptation of plants arrived as a consequence of natural or forced selection over many generations and it is based on genome modifications. On the other hand, acclimatisation of plants to a stressful environment considers morphophysiological changes and is mainly associated with phenotypic plasticity. Plants are acclimated to various conditions and respond flexibly to changes in cell metabolism and physiological activities as a response to changing environmental conditions [2]. Stress, as the imbalance between multiple environmental factors, could negatively affect growth, photosynthesis rate, membrane integrity, and protein stability [3]. Plants react on stressful conditions firstly at the cellular level, by altering its physical and chemical properties—primary effects of stress. Thereafter, metabolism disruption, formation of cytotoxic products, and injuries to membranes and other cell structures present the secondary effects of stress. The existing damage could

http://dx.doi.org/10.5772/60990

lead to irreversible changes with lethal consequences. Abiotic stress, such as drought, salinity, extreme temperatures, chemical toxicity, and oxidative stress are serious threats to agriculture [3]. By 2050, increased salinisation of arable land is expected to have devastating global impacts, resulting in 50% of land loss. For these reasons, special attention is given to the interaction between stressful conditions and plants in agricultural production, because the majority of agricultural crops are not highly adapted to the regions in which they are cultivated. The extent of damage caused to agricultural crops by a combination of two different stresses emphasises the necessity for the development of crops and plants with enhanced tolerance to combinations of different abiotic stresses [4].

2. Reactive oxygen species and redox reactions as stress moderators

One of the most important topics when the general impact of stress on plants is considered is free radical reactions, i.e., the production of reactive oxygen species (ROS). Their role in plant physiology is complex, altering plant metabolism and reactions to a changing environment in the direction of either adaptability or irreversible damage. The presence of stress increases the internal entropy of a plant, shifting it closer to equilibrium. Along with increased entropy, the requirements for energy increase (with increased intensity of respiration—exergonic processes) lead to an oxidative burst. During the increased oxidation and photorespiration, the produced high-energy electrons are transferring to molecular oxygen (O_2), inducing further ROS generation [5]. They include singlet oxygen ($^1O_2$), hydrogen peroxide (H_2O_2) and the superoxide anion (O_2^-) and hydroxyl (OH) radicals [6]. Their production also requires or releases some quantities of energy. The voltage of an electrochemical cell during ROS production is directly related to the change of the Gibbs free energy (Figure 1) [7, 8].

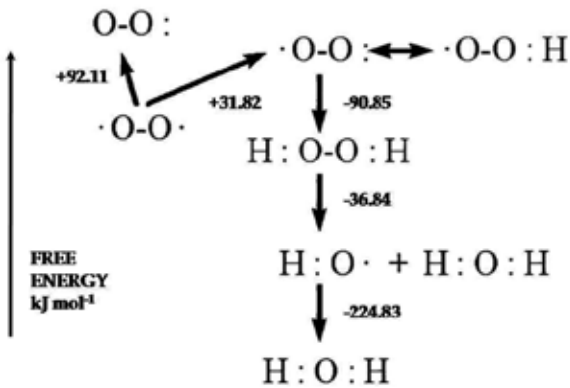

Figure 1. The free energy of different reactive oxygen species [7].

ROS production occurs mainly in chloroplasts and mitochondria (sites of electron transport) under abiotic stress, increasing the amounts of intracellular ROS to toxic levels. In this case,

plants activate several mechanisms to combat the increased ROS concentrations [9]. ROS, produced in both unstressed and especially in stressed plant cells, are also important signals as well as mediators in the biosynthesis of complex organic molecules, polymerisation of cell wall constituents and defence against various abiotic and biotic stresses [10]. Redox reactions provide electron flow over cascades of redox couples, where the reduction potential (reducing capacity) represents voltage, i.e., the number of available electrons. Redox reactions are responsible for energy production.

Thermodynamics (redox potential of oxidisable thiols) and kinetics (the ability to compete with the antioxidative system) of thiols are the key factors in evaluating the functional importance of thiol-based ROS sensors in plants [11]. Moreover, the signalling role of ROS is present in gene expression in response to high light, stomatal closure, responses that involve auxin, cell cycle, growth, and development [12]. From this viewpoint, ROS could act as an activator of hormones, as intermediatories between different signalling pathways.

Various abiotic stresses lead to ROS overproduction, which, because of their high reactivity, leads to proteins, lipids, carbohydrates, and DNA degradation. To combat the high levels of ROS in plant cells, plants possess very efficient enzymatic (superoxide dismutase, SOD; catalase, CAT; ascorbate peroxidase, APX; glutathione reductase, GR; monodehydroascorbate reductase, MDHAR; dehydroascorbate reductase, DHAR; glutathione peroxidase, GPX; guaicol peroxidase, GOPX; and glutathione-*S*-transferase, GST) and nonenzymatic (ascorbic acid; thiolic proteins, PSH; glutathione, GSH; phenolic compounds, alkaloids, nonprotein amino acids, carotenoides, and α-tocopherols) antioxidant systems of defence, acting mutually in order to protect plant cells from oxidative damage [13].

Plants have developed mechanisms to avoid ROS production: anatomical adaptations, such as leaf movement and curling, development of a refracting epidermis and the hiding of stomata in specialized structures; physiological adaptations such as C_4 and CAM metabolism; and molecular mechanisms that rearrange the photosynthetic apparatus and its antennae in accordance with the light quality and intensity, or completely suppress photosynthesis [12]. ROS production could also be decreased by the alternative channelling of electrons in the electron transport chains of the chloroplasts and mitochondria by a group of enzymes called alternative oxidases. Furthermore, genetic engineering developed strategies for cloning stress tolerance genes. For instance, a stress-activated novel aldose–aldehyde reductase cloned from alfalfa was inserted into tobacco plants, where the ectopic expression of this gene resulted in tolerance to oxidative stress and dehydration [14].

2.1. Light and radiation stress

Light is an energetic promoter of photosynthesis and one of the most important environmental inductors of development through the phytochrome system. Light is an environmental factor that can vary several times within a short period. The photosynthetic apparatus is liable to functional deactivation in high-intensity light, i.e., photoinhibition [15]. These processes are very important from the agricultural viewpoint, for biomass and yield production. Common afternoon photoinhibition (depression of photosynthesis during the highest daily insolation) reduces, on a daily basis, biomass production by about 8%–10% [16]. These processes are much

http://dx.doi.org/10.5772/60990

higher under stressful conditions, including intensive light. The photosynthetic apparatus must be able to maintain the requirements for maximal photon absorption pitted against the danger of excessive chlorophyll excitation under bright lights. In the shade, plants employ multiple cellular traits to maximize the efficient interception, absorption, and utilisation of light. However, under intensive light, they are challenged by a surplus of photo-energy, and if the light-driven electron transport exceeds the chloroplastic capacity to utilise this chemical reducing power, it could lead to ROS production (photo-oxidative stress) [17]. Prolonged illumination at saturated light intensities increases photodeactivation, from which plants will not able to recover or could only partially recover [18]. An important role was given to reduced tocopherol of the thylakoid membrane as a limiting factor for defensive reactions.

An optimal functioning of photosynthesis is maintained by coupling with other processes [19], such as (1) limited light absorption by chlorophyll, (2) dissipation of excess absorbed light as heat, (3) redirection of light-driven electron transport away from the energy-saturated Calvin cycle by alternative pathways, (4) lower number of functional photosystem II (PSII) centres by the delaying of chloroplast electron transport, and (5) maintaining of high antioxidant levels in the chloroplast complement to scavenge excess ROS [17]. However, these protective processes have the effect of lowering the energetic efficiency of photosynthesis. The other adaptation mechanisms include changes in leaf orientation or their rolling [15, 20], but the most efficient mechanism is available in the xantophyll cycle [21]. During photoinhibition induced by high light intensity, the concentration of zeaxanthin increases [22, 23]. It was also noticed that anthocyanins accumulate in the illuminated leaf surface of some maize genotypes at suboptimal temperatures, and xanthophyll de-epoxidation and SOD activities were lower in anthocyanin-containing than in anthocyanin-deficient maize [24]. Moreover, thiol regulation of the enzyme activities of the Benson–Calvin cycle link light-dependent electron pressure in the photosynthetic light reactions to ATP and NADPH consumption in reductive carbohydrate metabolism [25]. One of the protective mechanisms includes photobiosynthesis of isoprene and its release from the leaves into the environment as a dissipation of excess energy (entropy) [26]. This process terminates the sustained passage of thermodynamic flows and regulates the overall stability of the cell and the whole organism.

Radiation stress

An increase in ultraviolet-B (UV-B: 280–320 nm) radiation in lower atmospheric layers was present owing to the weakening of the stratospheric ozone layer, affecting plant growth and development [27]. Similar to photo-oxidative stress, radiation stress is mainly based on ROS production, which distresses a number of important physiological processes, such as photosynthesis and causes chemical and structural damage to DNA. This kind of stress induces more severe membrane damage than drought stress, as assessed by lipid peroxidation as well as osmolyte leakage [28]. In some cases, it could affect the integrity of plant genome eventually leading to cell death. From this point, plants response to UV stress by an increase in antioxidant synthesis, like anthocyanin and phenols, as well as osmolyte proline, such as that found in pea and wheat [28]. Induction of pathogenesis-related proteins as a defence mechanism is also mediated at the gene expression level [27]. Moreover, several DNA repair pathways are included in response to oxidative stress [29].

2.2. Low temperatures

Low temperatures slow down cellular metabolism, solute flow and growth, with increasing energy consumption. Their influence is particularly harmful during the reproductive developmental stage, when flower abscission, pollen sterility, pollen tube distortion, ovule abortion, and fruit set were reduced, ultimately lowering the yield [30]. A change in membrane fluidity, based on phase separation of membrane lipids, is one of the immediate consequences during low temperature stress. Photosynthesis is affected due to photoinhibitory injury of PSI [31], while ROS contribute to membrane damage and protein denaturation [30].

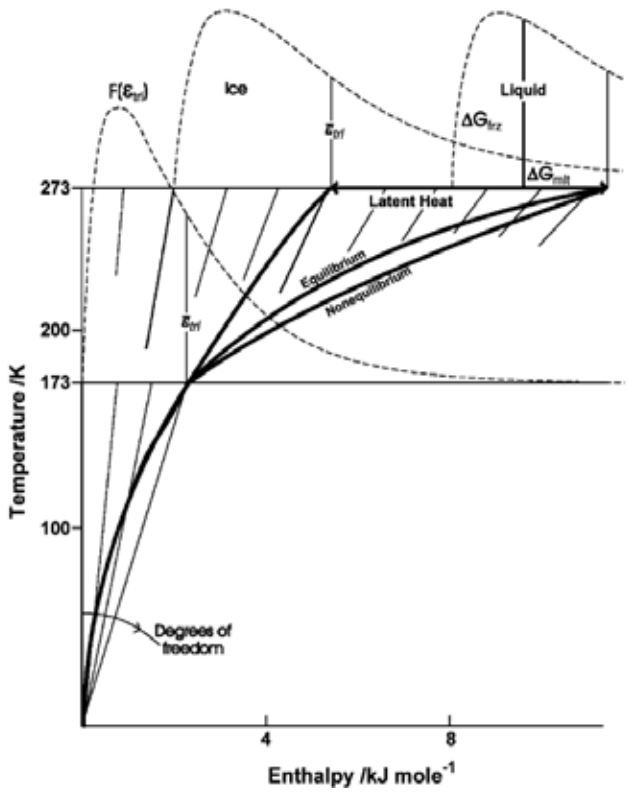

Figure 2. Enthalpy of ice (from 0 to 273 K), liquid water (at 273 K), and the empirical latent heat (from 248 to 273 K). The empirical values for the heat capacity of ice were used to calculate the energy from 0 to 173 K. The curve labelled as "$F(\varepsilon_{trl})$" is the frequency distribution of kinetic energy at 173 K. The line labeled ε_{trl} is the mean translational kinetic energy of this distribution. The lines indicating degrees of freedom (DF) are shifted to the right between 173 and 273 K because of the acquisition of nearly two additional unspecified DF, which are acquired by water as it melts. Note the convergence at 173 K of (1) the energy resulting from 3 degrees of freedom, (2) the integration of specific heat capacity, and (3) the latent heat for equilibrium and nonequilibrium freezing [34].

Chloroplasts are the first and the most severely affected organelles by chill [32]. The thylakoids swell and distort, starch granules disappear, and a peripheral reticulum appears. Chilling also leads to cytoplasmic acidification in chill-sensitive plants [38]. As the temperature drops below 0°C, the difference in concentration between intracellular (symplast) and extracellular fluids

http://dx.doi.org/10.5772/60990

(apoplast) causes the formation of intercellular ice crystals. Since the chemical potential of ice is lower compared to liquid water at a given temperature, unfrozen water moves toward the chemical potential gradient, from the symplast to the apoplast, causing cellular dehydration. Seeds and pollen, which are already in the state of anhydrobiosis, are more tolerant to ice formation. Initiation of nonequilibrium freezing with sufficient free energy that drives disruption exists in a supercooled system [34]. As the temperature continues to decrease, adhesion contributes to the disruptive effects at hydrated interfaces and protoplasts contract by freeze-dehydration. At 0°C, ice and liquid water are in a steady state of melting and freezing. A further temperature drop (i.e., <0°C), induces that freezing exceeds melting, free water freezes and an interface develops between crystals of nonmatching patterns (Figure 2). The existing adhesion increases at the expense of kinetic freedom, and causes a decrease in the latent heat. When all the free water is frozen, the enantiomorph becomes a simpler system of ice–interface–ice.

Cold acclimation of plants include pre-existing macromolecules and structural enzymes, structural proteins, lipids, and membranes, alterations in lipid composition, the appearance of new isozymes, increased levels of sugar, soluble proteins, proline, and organic acids [35]. Enhancement of antioxidative mechanisms, increased levels of sugars in the apoplastic space, and the induction of genes encoding molecular chaperones play important roles in the reduction of freezing-induced cellular damage [30]. Genotypes tolerant to low temperatures have the ability to change the thermodynamic properties of their membranes through phase transition from a flexible liquid-crystalline structure to a solid gel structure. Cold acclimation is thus connected to increases in the unsaturation levels of fatty acids [36], together with increased ratios of free sterols/glycolipids [37], as well as to the ability to synthesize antifreeze proteins (AFPs) [38].

2.3. High-temperature stress

In light of global climatic changes, heat stress and its combination with water deficits known as drought, are the most important abiotic stresses that affect crop growth, development, and yield. Temperature controls the rate of plant metabolism, consequently influencing the production of biomass, fruits and grains. The productivity of important agricultural crops is dramatically reduced when they experience short episodes of high temperatures during the reproductive period [39]. Heat stress is an important threat to global food supply, causing substantial crop yield losses.

Plant species originating from different areas have different temperature requirements for heath, as a factor for optimal growth, developmental stages, and yielding potential. Owing to the fact that rates of growth, cell division, and progression in the plant cycle are driven by temperature, following common Arrhenius-type response curves, a model of the thermal adaptation of contrasting genotypes grown in common ranges of temperatures was described (Figure 3) [40]. Species originating from cooler areas, such as canola and cauliflower, showed the lowest values for T_{opt}, while species adapted to warm climates, such as pearl millet, peanut, cotton, sorghum and cowpea, showed high values of T_{opt}.

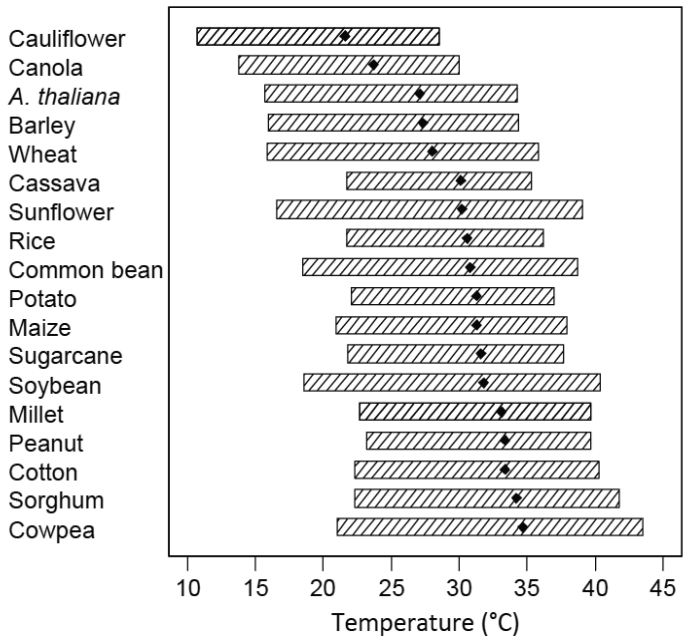

Figure 3. Temperature at which rates are at a maximum (T_{opt}, black dots) and range of temperature for which the rate of development is at least 50% of its maximum (horizontal bars) in 18 species [40].

Direct injuries in plants due to high temperatures include protein denaturation and aggregation, and increased fluidity of membrane lipids. Indirect injuries include deactivation of enzymes in chloroplast and mitochondria, inhibition of protein synthesis, protein degradation, and loss of membrane integrity [30]. As a consequence, disrupted integrity and functions of biological membranes enhance their permeability causing increased loss of electrolytes. In field experiments, at 40°C, significant photosynthetic losses arise from increased photorespiration (approximately about 40% of photosynthesis). Respiration requires greater carbon fixation for sustained growth and survival, while stomatal conductance, chlorophyll content and net photosynthetic rate are decreased [41]. One of the most important impacts of high temperature stress on cereals is the shortening of the developmental phase [42]. Furthermore, the synthesis of normal proteins and the accelerated transcription and translation of heat shock proteins (HSPs), together with the production of phytohormones, such as abscisic acid (ABA) and antioxidants, and changes in the organization of cellular structures, including organelles and the cytoskeleton, as well as membrane functions are present in response to stress [43]. Heat stress also leads to the oligomerisation of thioredoxins [44]. This reflects alterations in hormonal and redox homeostasis. In parallel with suppressed growth and accelerated developmental stages, other important consequences of heat shock are floral abortion and sterility, decreased pollen viability and grain filling, and decreased quality of the produced grains [45].

Thermotolerant genotypes have less depressed photosynthesis rates, with faster recovery after stress [46]. Maintenance of activity and increased transcription levels of antioxidant enzymes and nonenzymatic antioxidants, as well as photosynthesis are associated with variable

http://dx.doi.org/10.5772/60990

thermostability in heat-acclimatised plants, such as *Dactylis glomerata* L. [47]. Based on enthalpy conversion efficiency and ratio of oxidative phosphorylation to oxygen consumption (P/O ratio), genotypes tolerant to high temperatures are able to minimize entropy production and maximize the efficiency of mitochondrial energy conversions during stress while maintaining adequate finite rates of energy processing [48].

2.4. Water deficiency and drought stresses

When considering water stress, a different energy distribution in the plant is involved: energy could be consumed to maintain turgor by solute accumulation, on the growth of nonphotosynthetic organs, such as roots, to increase the capacity for water uptake, or on building the xylem elements, capable of surviving negative pressures. Plant responses to water scarcity are complex, and they could be synergistically or antagonistically modified by the presence of other stresses [49]. Different crop species could react in specific ways, such as changes in the root/shoot ratio, or the temporary accumulation of reserves in the stem, which are accompanied by alterations in nitrogen and carbon metabolism. The increased energy consumption from stored substances (carbohydrates and/or fatty acids) during drought is associated with increased expression of the enzymes involved in anabolic pathways, which respond with an increased content of some amino acids [50]. In the leaves, the dissipation of excitation energy through processes other than photosynthesis is also an important defence mechanism. For instance, in C_3 plants, energy dissipation by photosynthesis decreases under a mild drought, while dissipation by photorespiration increases in a compensatory manner. During moderate to severe drought, the contributions of both photosynthesis and photorespiration decrease, and thermal dissipation increases, consuming up to 70%–90% of the total light absorbed [51].

Accordingly, the negative impact of water deficit is reflected on photosynthesis decrease, along with decreases in stomatal conductance, viable leaf area, shoot and grain mass, as well as weight and soluble sugar content such as that found in wheat kernels [52]. On the other hand, the positive impact of drought on grain composition is reflected in the increased isoflavone content [53]. Whereas crops' response to drought considers many different responses, water lack–induced shrinkage of cell volume makes the cells more viscous, resulting in aggregation and denaturation of proteins and thus hindering the normal functioning of enzymes involved in photosynthesis, which, together with limited CO_2 influx, simultaneously enhances oxygenation, thereby increasing photorespiratory losses and ROS production [6]. Dramatic ROS increase causes damage by increasing lipid peroxidation, protein degradation, DNA fragmentation, and ultimately, cell death [54]. Decreased photosynthesis and poor translocation of assimilates from leaves are a consequence of highly decreased water potential in the phloem. The photosynthetic efficiency mainly depends on the openness of stomata, particularly in C_3 crops, while their closure tends to avoid excessive water loss. Abscisic acid (ABA) mediates water loss from the guardian cells of the stomata, which is triggered by a decrease in the water content of the leaf and inhibits leaf expansion. In muskmelon seedlings, ABA could improve the maintenance of the leaf water potential and relative water content, and reduce electrolyte leakage [55].

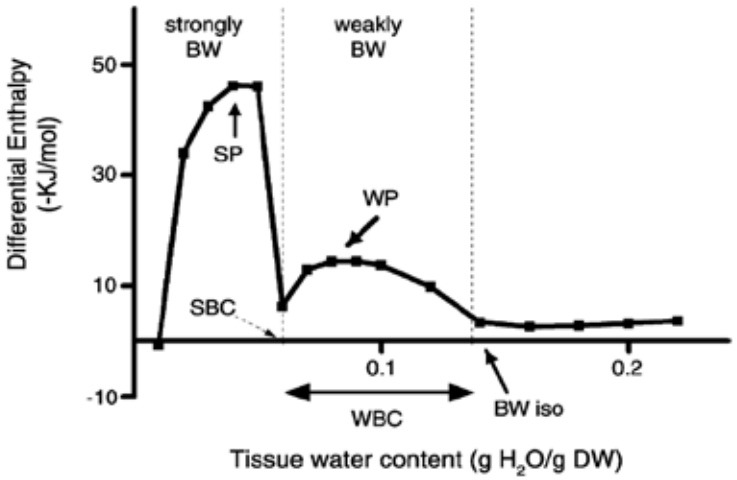

Figure 4. Examples of the differential sorption heat calculated from the sorption data using the Clausius–Clapeyron equation. The arrows depict the most negative enthalpy value in the region of strongly (SP) and weakly (WP) bound water, the limit of the strongly bound water region (SBC), the moisture content where bound water first appeared (BWiso), and the tissue moisture range corresponding to weakly bound water (WBC) [56].

The effects of drought, i.e., the quantitative properties of water in fresh and dry leaves of durum wheat were tested by the relation between the water status and the properties of bound water (BW) with different strengths to ionic, polar, or hydrophobic sites of macromolecules [56]. An increase in tissue affinity for strongly bound water implied a simultaneous increase in the affinity for weakly bound water. The qualitative properties of bound water may be particularly important for drought adaptation in durum wheat, which is associated with solute potential plots of differential energies of water sorption (Figure 4).

According to a systemic investigation of the impact of drought on cultivated plants, it was assumed that their response is complex. It mainly depends on drought duration, lasting from several hours to several days (short-term) and up to several weeks to months (long-term). The impact of water deficit on C metabolism differs among plant species, but the common characteristic for all plant species is that the C demand (growth) always decays before the C supply (photosynthesis) is affected by water deficit [57].

Signalling pathways in cultivated plants, in response to drought stress, enables the activation of a whole range of protective mechanisms. The catalase activity and the carotenoid and proline contents were increased in sunflower affected by drought [58]. A positive correlation between enzymatic and nonenzymatic antioxidants is present. Proline plays a significant role in drought tolerance, since tolerant wheat and maize genotypes, with a higher relative water content (RWC = (fresh weight – dry weight) / (turgid weight – dry weight)), had an overaccumulation of free proline, and a lower drought susceptibility index in comparison to drought-sensitive genotypes [59]. Stress tolerance includes the synthesis of heat shock proteins, which stabilize proteins and membranes, and support protein refolding and thus cellular homeostasis [4].

http://dx.doi.org/10.5772/60990

Further experiments should include the determination of the genetic basis of drought tolerance. Quantitative trait loci (QTLs) for drought-tolerant upland rice were determined, having positive effects on the net photosynthetic rate, stomatal conductance, transpiration rate, and the quantum yield of PS II [60]. The same allele is responsible for the improvement of rice grain yield and plant water uptake under drought conditions [61]. Also, a positive connection between improved water uptake and root architecture was found [62]. These studies could also include wild relatives or ancestors of common crops, which display high adaptability to different abiotic stresses, such as emmer wheat (*Triticum turgidum* ssp. dicoccoides), the ancestor of domesticated durum wheat (*Triticum turgidum* ssp. *durum*). Emmer wheat has microRNAs (miRNAs), which are a class of gene expression regulators that have also been linked to several plant stress responses [63]. In parallel, the direction of genetically modified crops was evolved. For instance, the trehalose-6-phosphate synthase gene from *Saccharomyces cerevisiae* was introduced in potato, which showed significantly increased drought resistance [64].

2.5. Water sufficiency, hypoxia, and anoxia

Water excess is a result of flooding, or soil concretion, or any other reason that could induce anaerobic conditions. Such conditions reflect on O_2 decrease (hypoxia) and, in some cases, to O_2 absence (anoxia). Roots react to such conditions by a partial loss of ability to conduct water to the shoot, leading to dehydration of the shoot [65]. Water logging forces the root to obtain most of the O_2 from the shoot through intervening tissues. Such a condition decreases root respiration and energy consumption increases, having as a consequence, imbalances of the energy potential in roots and shoots. Alternative metabolic pathways are activating, such as alcoholic fermentation, as well as several diverse fermentative bypasses, which ameliorate the poisoning through excessive accumulation of specific metabolic intermediates. In parallel to O_2 starvation, the roots are restrained from providing mineral nutrients for both themselves and the shoots. At the shoot level, CO_2 incorporation is depleted owing to stomatal closure [66]. As a consequence of the general reduction in metabolic activity, a significant reduction in the net photosynthetic rate, chlorophyll a and b contents (with an increase in chlorophyll *a*/*b* ratio), and in the efficiency of water use and intrinsic water use were reported [67]. Furthermore, the decreased leaf water potentials cause visible wilting, having as a consequence an increase in the ABA concentration in the shoots. Different crop species have specific responses to hypoxia. In wheat, narrow-leafed lupin and yellow lupin hypoxia affected solute transport, increasing the root pressure (Pr) and decreasing the turgor pressure (Pc), but only significantly in lupin. Different pathways for radial water flow across the roots of lupin and wheat were observable, with increased aquaporin activity in wheat roots [68].

The existence of energy and carbohydrate crises during hypoxia has to be controlled by regulated consumption of carbohydrates and energy reserves [69]. From this viewpoint, in tolerant species, such as rice, several adaptation strategies exist: in seeds germinated under water, energy homeostasis and growth are connected by a calcineurin B-like interacting binding kinase. At the shoot level, two opposing adaptive strategies are present, i.e., elongation (escape) and inhibition of elongation (quiescence), which are controlled by related ethylene

response factor DNA binding proteins that act downstream of ethylene and modulate gibberellin-mediated shoot growth. Hypoxia, as many other abiotic stresses, induces ROS formation [70].

Efficient utilization of energy resources (starch, sugars), together with a switch to anaerobic metabolism and preservation of the redox status of the cell are vital for survival during hypoxia. Plants can escape hypoxia stress through multifaceted alterations at the cellular and organ levels, and by structural changes that promote access to and diffusion of O_2. These processes are driven by phytohormones, including ethylene, gibberellins (GA), and ABA [70]. The early increase in cytosolic Ca^{2+}, as well as the rapid establishment of ionic homeostasis, may be essential for the induction of adaptive changes at the cellular and organismal levels during anaerobiosis [71]. This could be, to a greater extent, connected to altered metabolism, firstly by increased activities of fermentative and glycolytic enzymes, while in hypoxia-tolerant rice, only minor metabolic activity was observed [72]. Experiments with the reactions of intolerant and tolerant crops to anoxia showed that sensitive ones quickly lose viability, with a strong metabolic arrest of sucrolytic, glycolytic, and fermentative enzymes. However, rice is able to keep the ATP level at 25% of the level found under aerated conditions. According to the differences observed in tolerance mechanisms against the effects of water logging, they were grouped into adaptive traits in relation to: (1) phenology, (2) morphology and anatomy, (3) nutrition, (4) metabolism including anaerobic catabolism and anoxia tolerance, and (5) post-anoxic damage and recovery [73]. The best opportunities for germplasm improvement were found in further utilization of genetic diversity, including the use of marker-assisted selection.

2.6. Imbalance in soil minerals, salinity, and acidity stress

Imbalances in the soil minerals could affect plant growth and development, by affecting the nutritional status of the plant, or through water uptake, or through toxic effects of ions on plant cells. Lack in important plant mineral nutrients (N, P, K, Ca, Mg, Fe, Zn, Mn, Cu, B, or Mo) could affect many physiological processes. Conversely, sufficiency in some mineral elements could restrain the availability of other ones, or could be toxic to cultivated plants. Plants have evolved adaptive mechanisms that enable the uptake of most mineral nutrients by the rhizosphere using root exudates. They consist of a complex mixture of organic acid anions, phytosiderophores, sugars, vitamins, amino acids, purines, nucleosides, inorganic ions (e.g., HCO_3^-, OH^-, H^+), gaseous molecules (CO_2, H_2), and enzymes that have direct or indirect effects on the acquisition of mineral nutrients required for plant growth [74].

High concentrations of salts (e.g., Na, Cl, and other ions), i.e., salinity stress, have two modes of action, i.e., nonspecific osmotic stress caused by water deficit and specific ion effect resulting from the accumulation of Na and Cl ions, which disturb nutrient acquisition and induce cytotoxicity. Salinity could have several origins: soils which lay on geologic marine deposits, proximity of a seashore, and improper water and fertilization management. It is well known that transpiration and evaporation take away water from the soil as vapour, concentrating the minerals in the soil solution.

Decreased plant growth caused by salinity is divided into two phases: the initial phase is due to an osmotic effect (it is similar to the initial response to water stress), while the second, slower

phase is the result of salt toxicity in leaves [75]. The damaging effects of salinity to sensitive plant species, such as soybean and cotton, are observable through a lack of germination, plant growth, decreased root growth, shoot and leaf biomass decrease, as well as in increased Na^+ and Cl^- concentrations in the leaves [76, 77]. Salinity also decreased the Ca^{2+} and Mg^{2+} concentrations in leaves, as well as the water potential (ψ_w) and solute potential (ψ_π) in quinoa seedlings [78]. The differential enthalpy varied significantly due to the presence of different ions in the solute. Such results indicate possible domination of exothermic reactions. In more tolerant wheat genotypes, increase in the contents of total water-soluble carbohydrates, glucose, fructose, sucrose, and fructans were evidenced under osmotic stress [79], indicating a higher availability of energetic substances and, thus, better growth potential.

One of the most important tolerance mechanisms which suppress Na^+ toxicity is the sequestration of excess Na^+ in the vacuole by vacuolar Na^+/H^+ pump using a pH gradient generated by H^+-ATPase and H^+-pyrophosphatase to maintain a higher K^+/Na^+ ratio in the cytoplasm [80]. The important role of Ca must not be forgotten, since it maintains plasma membrane selectivity for K^+ over Na^+ [75]. Together with ion homeostasis, it is important to preserve redox homeostasis during salinity stress [81]. Many tolerance mechanisms are included in the restraining of salinity stress. The role of osmolytes (K^+ and organic solutes) and different osmoprotectants: carbohydrates, soluble proteins, late-embryogenesis abundant (LEA) proteins, amino acids (proline is of high significance), asparagine, quaternary ammonium compounds (glycine betaine, β-alanine betaine, and proline betaine), and polyols (such as mannitol, glycerol, sorbitol, ononitol, and pinitol) was emphasized [82].

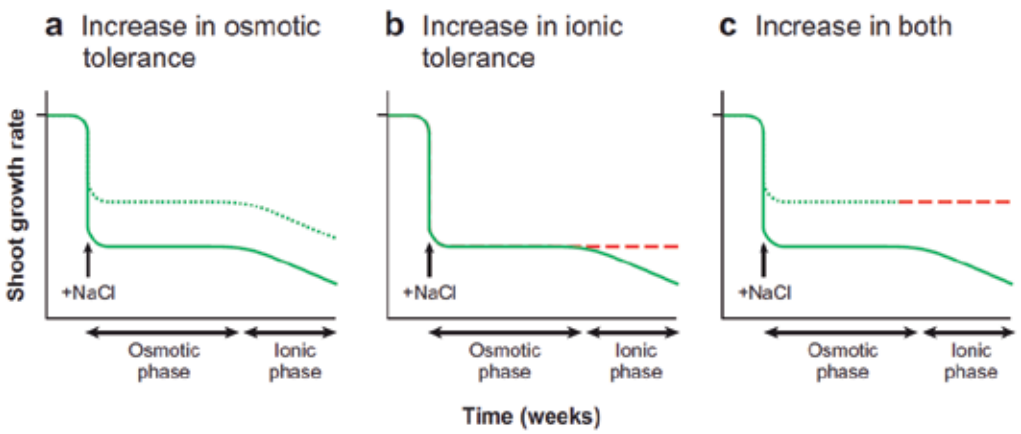

Figure 5. The growth response to salinity stress. The solid green line represents the change in the growth rate after the addition of NaCl. (*a*) The broken green line represents the hypothetical response of a plant with an increased tolerance to the osmotic component of stress. (*b*) The broken red line represents the response of a plant with an increased tolerance to the ionic component of stress. (*c*) The green-and-red line represents the response of a plant with increased tolerance to the osmotic and ionic components of stress [83].

Three distinct types of plant adaptations to salinity were determined: osmotic stress tolerance, Na^+ or Cl^- exclusion, and the tolerance of the tissue to accumulated Na^+ or Cl^- (Figure 5) [83].

The first phase is associated with the activation of osmolytic and osmoprotective substances. In the second, ion-specific phase, the salt accumulates to toxic concentrations in the older leaves which do not grow and hence lose their ability to dilute salt. They die and when the dying rate is faster than the production of new leaves, the growth rate is decreased (owing to a reduced photosynthetic capacity). This signifies that osmotic stress has a greater impact on growth than ionic stress. The effect of increased tolerance to osmotic stress is shown in Figure 5a; an increase in ionic tolerance is presented in Figure 5b, while the combined tolerance is shown in Figure 5c. The adaptability to salinity also includes Na^+ exclusion from leaves by roots according to the difference in chemical potential between cytosolic and xylem Na^+ concentrations [83]. This mechanism also involves a Na^+/H^+ antiporter$^+$. There are also species that accumulate high Cl^- concentrations in leaves, such as soybean, avocado and species that have Cl^--excluding rhizomes (such as grapevines and citrus), for which Cl^- toxicity is more important than Na^+ toxicity.

In parallel with salinity, soil acidity has harmful effects on plants by the decreased pH value and the increased levels of Al, Fe, and Mn ions. It is associated with corresponding deficiencies in available P, Mo, Ca, Mg, and K [84]. Soil acidity is a limiting factor that affects the growth, height, and yield stability of many crops worldwide. It is a consequence of pedogenesis factors, intensive application of higher amounts of mineral fertilizers with low pH, low inputs of organic fertilizers, and acid rain. Acidity affects about 4 billion ha, representing 30% of the total ice-free land area of the world [85].

Low pH, expressed through proton (H^+) toxicity, was mainly expressed as the inhibition of root elongation and root death, varying between different plant species and genotypes within the same species [86]. This type of toxicity has three modes of action: (1) disruption of cell wall integrity, (2) disturbance in cytosolic pH stability, and (3) inhibition in the uptake of cations. A high H^+ level interrupts the polysaccharide network in cell walls by displacing Ca^{2+}, together with impairment of the plasma membrane H^+-ATPase to maintain the cytosolic pH. The third mode is associated with depolarisation of the plasma membrane, thus becoming incapable for cation uptake. At low pH, and in parallel with H^+ toxicity, Al^{3+} is considered to be the most phytotoxic Al form: it inhibits the root growth of maize through binding to sensitive binding sites in the apoplast of the epidermis and the outer cortex, while $Al(OH)_3$ precipitation causes a mechanical barrier. Al toxicity leads to inhibited cell elongation and cell division, which is accompanied by reduced water and nutrient uptake [87].

Soil acidity presents a stressful factor that can be controlled by adequate soil management. $CaCO_3$ and $CaSO_4$ incorporation reflects an increase of the pH values and a decrease in the concentrations of exchangeable H^+ and Al^{3+}, respectively. Simultaneously, the concentration of available P and K increases [86, 88]. Organic matter is an important factor in the amelioration of soil acidity with regard to Al, which forms complexes with organic ligands, particularly with organic acid anions. Humic acid deserves special attention as it shows buffer behaviour on base or acid addition, it expresses buffer action between pH 5.5 and 8.0 [89].

Not all species and genotypes of the same species show the same tolerance to growth on acid soils. This kind of tolerance is a complex phenomenon (it includes tolerance to H^+ and Al^{3+} toxicity), parallel to lower availability of nutrients (P, Mo, Ca, Mg, and K). Al tolerance is a

http://dx.doi.org/10.5772/60990

complex multigenic trait, associated with organic acid syntheses [87]. It is important to emphasise that some plants, such as rooibos tea (*Aspalathus linearis* L.), actively modify the pH of their rhizosphere by extruding OH^- and HCO_3^- in order to facilitate growth in low pH soils (pH 3–5) [74]. The degree of Al tolerance by transgenic alfalfa plants (with an inserted gene for the synthesis of the phosphoenolpyruvate carboxylase enzyme) supports the concept that enhancing organic acid synthesis in plants may be an effective strategy to cope with soil acidity and Al toxicity [90].

2.7. Pesticide and heavy metals (xenobiotics) stress

Some toxic substances (organic pollutants, pesticides, heavy metals, and other natural or synthetic toxins—xenobiotics), could induce stress in cultivated plants. This type of stress could also be connected with inadequate cropping management or pronounced plant sensitivity. Modern agricultural production is inconceivable without herbicide application to facilitate weed removal and substitution of destructive soil cultivation. Herbicides target specific enzymes, and thus many resistance-endowing mutations may occur in weeds, creating the need to investigate new formulations [91]. Crops could also be injured by low herbicide selectivity or increased susceptibility. It is particularly noticeable during the accompanied presence of herbicide and some other stress, such as high temperature, water deficit, etc. [92].

Herbicides can affect photosynthesis by decreasing the quantum yield, which, together with stomatal closure and decrease in CO_2 assimilation, affects energy metabolism [93]. Some herbicides induce photodamage [94]. Decreased photosynthesis efficiency reflects to an increased need for osmotic adjustment in affected tissues [95]. Herbicides could also induce ROS production [13], disturb redox homeostasis [96, 97], and affect protein metabolism [98] leading to drying and lack of growth. Phytotoxicity, in the case of herbicide stress, acts on energy disposal in plants through two processes [99]. After the rapid metabolisation of herbicide molecules, the second process includes recovery from damage caused by molecules that reached the action site of the herbicide. The energy required for recovery from phytotoxicity symptoms must be taken from other processes resulting in larger or smaller yield losses. Herbicide application might result in temporary or permanent stress, depending on the herbicide's characteristics (mode of application and rate), type of the crop (inbred line, cultivar, or hybrid), developmental stage, nutrition, water balance, and the environment. Temporary stress allows rapid plant recovery from damage (with later recovery to the initial growth rate), with lower or without yield losses, but with relevant changes in the crop cycle. On the other hand, permanent stress reduces the plant growth rate, in a way that the probability of yield losses is greater (Figure 6).

Herbicides could be absorbed rapidly by cultivated plants, and their effect on metabolism could be noticed a few days after application. Phytotoxic effects of different herbicides have diverse impacts on the energetic properties of plants. For instance, nicosulfuron increases energy consumption and foramsulfuron induces "metabolic burst" [100]. Phytotoxicity correlates with enthalpy increase (ΔH), indicating endothermic reactions [101]. This means that a great deal of the potential energy of the plant was metabolically consumed [102]. Weeds as competitors also disturb the energy efficiency in cultivated plants.

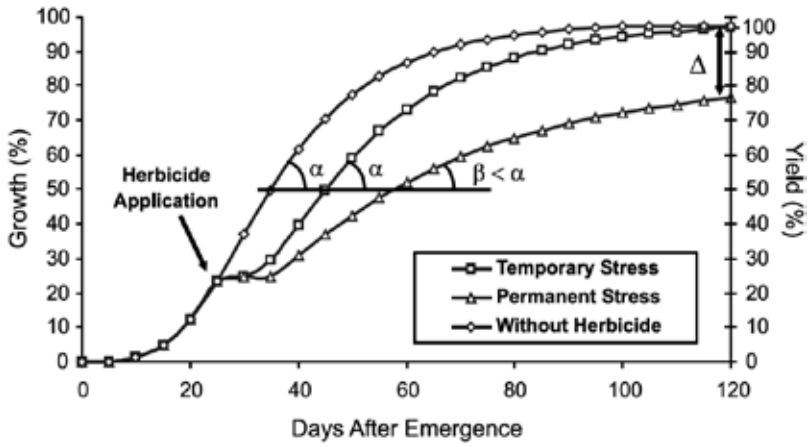

Figure 6. Mathematical models for the effects of herbicide phytotoxicity on crop growth and yield. Δ represents an estimation of final yield reduction [99].

Increased photorespiration and changes in carbon metabolism-associated proteins stimulate the synthesis of a set of pathogenesis-related proteins, suggesting that they could play an essential role in cell defence against herbicide stress [95]. An increase in the level of antioxidants such as phytate, phenolics, and thiolic proteins, as well as their protective activity, were also observed [96, 103]. The present genetic variability in herbicide tolerance is one of the most employed tactics to obtain herbicide-tolerant crops. This method is mainly based on naturally occurring variability or is the consequence of mutagenesis techniques. This mechanism was used in the production of cycloxydim-tolerant maize (CTM) mutation, which is the dominant gene that conferred tolerance to the herbicide Cycloxydim [104].

Heavy metals are also considered as dangerous xenobiotics. Some of them are essential for plant metabolism (such as Zn, Ni, and Cu), as they are important constituents of pigments and enzymes, while the others (such as Cd, Pb, Hg, and Cu) are toxic in higher concentrations. Plants could accumulate them in lower concentrations without disrupting their own metabolisms, but then becoming sources of heavy metal contamination in the food chain. This was confirmed by the positive correlation between Cd and Pb concentrations in soil and in the fruit of the asparagus bean [105]. In plants, heavy metals disrupt the functions of enzymes, replace essential metals in pigments, and induce the production of ROS and methylglyoxal, which could cause peroxidation of lipids, oxidation of proteins, deactivation of enzymes, DNA damage and/or interact with other vital constituents of plant cells [106, 107].

Plants could maintain the necessary concentrations of essential metal ions in cells by homeostatic mechanisms. They are also involved in the reduction of damage induced by heavy metal toxicity. One of the major tolerance mechanisms is chelation of heavy metals by a family of peptide ligands, the phytochelatins (similar to other xenobiotics) [108]. The molecular basis for the chelators and chaperone synthesis is well known and could be applied in the modification of tolerant plants. Tolerance to Cd and As is largely dependent on the phytochelatin pathway, but molecular biology of Cd hypertolerance in certain plant species, such as the

metallophytes *Arabidopsis halleri* or *Thlaspi caerulescens*, is still under study [109]. Heavy metals may be sequestered by amino acids, organic acids, and glutathione (GSH), which have a central role in ROS scavenging mechanisms, as well as in the glyoxalase system [110]. This means that GSH and its metabolising enzymes (glutathione S-transferase, glutathione peroxidase, dehydroascorbate reductase, glutathione reductase, glyoxalase I, and glyoxalase II) are involved in the direct and indirect control of ROS and methylglyoxal. All the facts mentioned above emphasise the importance of genotypes tolerant to heavy metal toxicity, i.e., those which accumulate lower concentrations of heavy metals or have developed strong detoxifying mechanisms. Some agricultural crops could be safely cultivated on soils polluted with heavy metals because of their low phytoextraction ability [105, 110].

Plants play a great role in the control of soils polluted with heavy metals through phytoremediation, which could be accomplished through [106]: (1) phytoextraction—accumulation of heavy metals from soils in plant organs that can be harvested; (2) rhizofiltration—decontamination of polluted waters and sewage by absorbing or uptake by the plant roots; (3) phytodegradation—utilization of the ability of some plants to decompose (degrade) pollutants; (4) phytostabilization—storage of heavy metals or other pollutants in plant tissues in the form of sparingly soluble complexes; and (5) phytovolatilisation—detoxification of soils by plants with the ability to produce volatile compounds.

GSH and many other molecules with various functional groups in plant cells (e.g., carboxyl, amine, hydroxyl, and sulphydryl), or some organic molecules (e.g., lignin, chitin, and humic substances) have sorption capabilities for heavy metals, i.e., they have biosorbent capacity. Various waste materials such as fly ash, slag, red mud, water treatment sludge, fungal and bacterial biomass, tree bark, sawdust, paper mill sludge, seafood processing wastes, and composted organics also have such properties [111] and could be used for remediation of polluted soils. The necessity to develop low-cost sorbents with a wide range of metal affinities (through the combination of several waste sorbent materials) that could remove a variety of metal ions from multielement-contaminated waters or soils as a remediation practice was emphasized.

3. Conclusion

Variations in environmental factors shift plant metabolism out of homeostasis. In the attempt to regain homeostasis, plants spend lower or higher amounts of energy. For increased yield (grain or biomass) cultivated plants require additional energy for successful completion of their life cycle, which makes them especially susceptible to stressful environments. In general, the stress suppresses many physiological processes, from photosynthesis, respiration, water absorption and its flow, up to hormonal and redox balance. Except for soil acidity and xenobiotics, which can be controlled by adequate agricultural management, all other stress factors are still beyond control.

Plants are forced to utilize stored and/or additional energy to combat stressful conditions, reducing yield potential temporarily during the recovery process, or permanently with

irreversible changes (significant yield drop). Mutations and breeding could give the ability for selected genotypes to adapt and acclimatise to environmental variations, and thus enable them to reduce stress pressure through metabolic alterations or through the synthesis of secondary metabolites and other protective substances. However, such activities reduce the energetic potential of the plant to some extent, but the potential for the plants' survival is increased. From this viewpoint, the necessity for the development of tolerant genotypes, which require less energy for maintenance of homeostasis, arises. Agricultural practices, such as irrigation, fertilization, cultivation, and pesticide application, could also reduce the impact of stress on cultivated plants. However, the protective measures of these agricultural practices could be limited if the stress is severe or long-term and if the crop is susceptible. Accordingly, the best solution for stable and high yield could be achieved through a combination of a genotype potential to reach homeostasis and an agricultural practice that maintains the environmental impact as close to optimum as possible.

Acknowledgements

This work was supported by Projects TR 31068 and TR31037 from the Ministry of Education, Science and Technological Development, Republic of Serbia.

Author details

Vesna Dragičević

Address all correspondence to: vdragicevic@mrizp.rs

Department for Agroecology and Crop Technology, Maize Research Institute "Zemun Polje", Zemun Polje, Serbia

References

[1] Leshem YY., Kuiper PJC. Is there a gas (general adaptation syndrome) response to various types of environmental stress? Biologia Plantarum 1996;38(1) 1–18.

[2] Lichtenthaler HK. The stress concept in plants: an introduction. Annals of the New York Academy of Sciences 1998;851 187–198.

[3] Wang W., Vinocur B., Altman A. Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. Planta 2003;218(1) 1–14.

[4] Mittler R. Abiotic stress, the field environment and stress combination. Trends in Plant Science 2006;11(1) 15–19.

[5] Bhattacharjee S. Reactive oxygen species and oxidative burst: roles in stress, senescence and signal transduction in plants. Current Science 2005;89(7) 1113–1121.

[6] Impa SM., Nadaradjan S., Jagadish SVK. Drought Stress Induced Reactive Oxygen Species and Anti-oxidants in Plants. In: Ahmad P., Prasad MNV. (ed.) Abiotic Stress Responses in Plants: Metabolism, Productivity and Sustainability, Springer Science+ Business Media 2012, p. 131–147.

[7] Vitvitskii, AI. Activation energy of some free-radical exchange reactions. Theoretical and Experimental Chemistry 1969;5(3) 276–278.

[8] Buettner GR. The packing order of free radicals and antioxidants: lipid peroxidation, α-tocopherol, and ascorbate. Archives of Biochemistry and Biophysics 1993;300(2) 535–543.

[9] Apel K., Hirt H. Reactive oxygen species: metabolism, oxidative stress, and signal transduction. Annual Review of Plant Biology 2004;55 373–399.

[10] Dat J., Vandenabeele S., Vranová E., Van Montagu M., Inzé D., Van Breusegem F. Dual action of the active oxygen species during plant stress responses. Cellular and Molecular Life Sciences CMLS 2000;57(5) 779–795.

[11] Foyer CH. Redox homeostasis and antioxidant signaling: a metabolic interface between stress perception and physiological responses. The Plant Cell 2005;17(7) 1866–1875.

[12] Mittler R. Oxidative stress, antioxidants and stress tolerance. Trends in Plant Science 2002;7(9) 405–410.

[13] Gill SS., Tuteja N. Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. Plant Physiology and Biochemistry 2010;48(12) 909–930.

[14] Bartels D. Targeting detoxification pathways: an efficient approach to obtain plants with multiple stress tolerance? Trends in Plant Science 2001;6(7) 284–286.

[15] Larcher W. Solar radiation Stress In: Physiological Plant Ecology. Ecophysiology and Stress Physiology of Functional Groups. 4th Edition, Berlin, Heidelberg and New York, Springer-Verlag 2003, p. 357–364.

[16] Werner C., Ryel RJ., Correia O., Beyschlag W. Effects of photoinhibition on whole-plant carbon gain assesed with a photosynthesis model. Plant Cell and Environment 2001;24 27–40.

[17] Skillman JB., Griffin KL., Earll S., Kusama M. Photosynthetic productivity: can plants do better? In: Piraján JCM. (ed.) Thermodynamics—Systems in Equilibrium and

Non-Equilibrium. Rijeka, InTech 2011. p. 35–68. Available from: http://cdn.intechopen.com/pdfs-wm/21498.pdf.

[18] Graßes T., Grimm B., Koroleva O., Jahns P. Loss of α-tocopherol in tobacco plants with decreased geranylgeranyl reductase activity does not modify photosynthesis in optimal growth conditions but increases sensitivity to high-light stress. Planta 2001;213(4) 620–628.

[19] Noctor G., Foyer CH. A re-evaluation of the ATP:NADPH budget during C_3 photosynthesis. A contribution from nitrate assimilation and its associated respiratory activity? Journal of Experimental Botany 1998;49 1895–1908.

[20] Sarieva GE., Kenzhebaeva SS., Lichtenthaler HK. Adaptation potential of photosynthesis in wheat cultivars with a capability of leaf rolling under high temperature conditions. Russian Journal of Plant Physiology 2010;57(1) 28–36.

[21] Müller P., Li XP., Niyogi KK. Non-photochemical quenching: a response to excess light energy. Plant Physiology 2001;125 1558–1566.

[22] Demmig-Adams B., Adams WW. III The role of xanthophylls cycle carotenoids in the protection of photosynthesis. Trends in Plant Science 1996;1 21–26.

[23] Horton P, Ruban AV., Walters RG. Regulation of light harvesting in green plants. Annual Review of Plant Physiology and Plant Molecular Biology 1996;47 65–84.

[24] Pietrini F., Iannelli MA., Massacci A. Anthocyanin accumulation in the illuminated surface of maize leaves enhances protection from photo-inhibitory risks at low temperature, without further limitation to photosynthesis. Plant, Cell & Environment 2002;25(10) 1251–1259.

[25] Dietz K-J., Pfannschmidt T. Novel regulators in photosynthetic redox control of plant metabolism and gene expression. Plant Physiology 2011;155(4) 1477–1485.

[26] Sanadze GA. Photobiosynthesis of isoprene as an example of leaf excretory function in the light of contemporary thermodynamics. Russian Journal of Plant Physiology 2010;57(1) 1–6.

[27] Mackerness SA-H. Plant responses to ultraviolet-B (UV-B: 280–320 nm) stress: What are the key regulators? Plant Growth Regulation 2000;32(1) 27–39.

[28] Alexieva V., Sergiev I., Mapelli S., Karanov E. The effect of drought and ultraviolet radiation on growth and stress markers in pea and wheat. Plant, Cell & Environment 2001;24(12) 1337–1344.

[29] Balestrazzi A., Confalonieri M., Macovei A., Donà M., Carbonera D. Genotoxic stress and DNA repair in plants: emerging functions and tools for improving crop productivity. Plant Cell Reports 2011;30(3) 287–295.

[30] Waraich E.A., Ahmad R., Halim A., Aziz T. Alleviation of temperature stress by nutrient management in crop plants: a review. Journal of Soil Science and Plant Nutrition 2012;12(2) 221–244.

[31] Yordanov I., Velikova V. Photoinhibition of photosystem 1. Bulgarian Journal of Plant Physiology 2000;26(1–2) 70–92.

[32] Kratsch HA., Wise RR. The ultrastructure of chilling stress. Plant, Cell & Environment 2000;23(4) 337–350.

[33] Kawamura Y. Chilling induces a decrease in pyrophosphate-dependent H^+-accumulation associated with a ΔpH_{vac}-stat in mung bean, a chill-sensitive plant. Plant, Cell & Environment 2008;31(3) 288–300.

[34] Olien CR., Livingston DP.III Understanding freeze stress in biological tissues: thermodynamics of interfacial water. Thermochimica Acta 2006;451(1–2) 52–56.

[35] Thomashow MF. Molecular Genetics of cold acclimation in higher plants. Advances in Genetics 1990;28 99–131.

[36] Kargiotidou A., Deli D., Galanopoulou D., Tsaftaris A., Farmaki T. Low temperature and light regulate delta 12 fatty acid desaturases (FAD2) at a transcriptional level in cotton (*Gossypium hirsutum*). Journal of Experimental Botany 2008;59(8) 2043–2056.

[37] Bohn M., Lüthje S., Sperling P., Heinz E., Dörffling K. Plasma membrane lipid alterations induced by cold acclimation and abscisic acid treatment of winter wheat seedlings differing in frost resistance. Journal of Plant Physiology 2007;164(2) 146–156.

[38] Xu H-N., Huang W., Jia C., Kim Y., Liu H. Evaluation of water holding capacity and breadmaking properties for frozen dough containing ice structuring proteins from winter wheat. Journal of Cereal Science 2009;49(2) 250–253.

[39] Teixeira EI., Fischer G., van Velthuizen H., Walter C., Ewert F. Global hot-spots of heat stress on agricultural crops due to climate change. Agricultural and Forest Meteorology 2013;170 206–215.

[40] Parent B., Tardieu F. Temperature responses of developmental processes have not been affected by breeding in different ecological areas for 17 crop species. New Phytologist 2012;194(3) 760–774.

[41] Morales D., Rodríguez P., Dell'Amico J., Nicolás E., Torrecillas A., Sánchez-Blanco MJ. High-temperature preconditioning and thermal shock imposition affects water relations, gas exchange and root hydraulic conductivity in tomato. Biologia Plantarum 2003;47(2) 203–208.

[42] Bonhomme R. Bases and limits to using 'degree.day' units. European Journal of Agronomy 2000;13(1) 1–10.

[43] Barnabás B., Jäger K., Fehér A. The effect of drought and heat stress on reproductive processes in cereals. Plant, Cell & Environment 2008;31(1) 11–38.

[44] Hemantaranjan A., Nishant Bhanu A., Singh MN., Yadav DK., Patel PK., Singh R., Katiyar D. Heat stress responses and thermotolerance. Advances in Plants Agriculture Research 2014;1(3) 00012.

[45] Prasad PVV., Boote KJ., Allen Jr. LH. Adverse high temperature effects on pollen viability, seed-set, seed yield and harvest index of grain-sorghum [*Sorghum bicolor* (L.) Moench] are more severe at elevated carbon dioxide due to higher tissue temperatures. Agricultural and Forest Meteorology 2006;139(3–4) 237–251.

[46] Zhao XX., Huang LK., Zhang XQ., Li Z., Peng Y. Effects of heat acclimation on photosynthesis, antioxidant enzyme activities, and gene expression in orchardgrass under heat stress. Molecules 2014;19 13564–13576.

[47] Macfarlane C., Adams MA., Hansen LD. Application of an enthalpy balance model of the relation between growth and respiration to temperature acclimation of *Eucalyptus globulus* seedlings. Proceedings of the Royal Society of London B 2002;269(1499) 1499–1507.

[48] Chaves MM., Pereira JS., Maroco J., Rodrigues ML., Ricardo CPP., Osório ML., Carvalho I., Faria T., Pinheiro C. How plants cope with water stress in the field? Photosynthesis and growth. Annals of Botany 2002;89(7) 907–916.

[49] Shu L., Lou Q., Ma C., Ding W., Zhou J., Wu J., Feng F., Lu X., Luo L., Guowang X., Mei H. Genetic, proteomic and metabolic analysis of the regulation of energy storage in rice seedlings in response to drought. Proteomics 2011;11(21) 4122–4138.

[50] Siddique MRB., Hamid A., Islam MS. Drought stress effects on water relations of wheat. Botanical Bulletin of Academia Sinica 2000;41 35–39.

[51] Shah NH., Paulsen GM. Interaction of drought and high temperature on photosynthesis and grain-filling of wheat. Plant and Soil 2003;257(1) 219–226.

[52] Krishnan P., Singh R., Verma APS., Joshi DK., Singh S. Changes in seed water status as characterized by NMR in developing soybean seed grown under moisture stress conditions. Biochemical and Biophysical Research Communications 2014;444 485–490.

[53] Caldwell CR., Britz SJ., Mirecki RM. Effect of temperature, elevated carbon dioxide, and drought during seed development on the isoflavone content of dwarf soybean [*Glycine max* (L.) Merrill] grown in controlled environments. Journal of Agricultural Food Chemistry 2005;53(4) 1125–1129.

[54] Anjum SA., Xie X-Y., Wang L-C., Saleem MF., Man C., Lei W. Morphological, physiological and biochemical responses of plants to drought stress. African Journal of Agricultural Research 2011;6(9) 2026–2032.

[55] Agehara S. Leskovar DI. Characterizing concentration effects of exogenous abscisic acid on gas exchange, water relations, and growth of muskmelon seedlings during

water stress and rehydration. Journal of the American Society for Horticultural Science 2012;137(6) 400–410.

[56] Rascio A., Nicastro G., Carlino E., Di Fonzo N. Differences for bound water content as estimated by pressure–volume and adsorption isotherm curves. Plant Science 2005;169(2) 395–401.

[57] Muller B., Pantin F., Genard M., Turc O., Freixes S., Piques M., Gibon Y. Water deficits uncouple growth from photosynthesis, increase C content, and modify the relationships between C and growth in sink organs. Journal of Experimental Botany 2011;62(6) 1715–1729.

[58] Ghobadi M., Taherabadi S., Ghobadi M-E., Mohammadi G-R., Jalali-Honarmand S. Antioxidant capacity, photosynthetic characteristics and water relations of sunflower (*Helianthus annuus* L.) cultivars in response to drought stress. Industrial Crops and Products 2013;50 29–38.

[59] Kravić N., Marković K., Anđelković V., Hadži-Tašković Šukalović V., Babić V., Vuletić M. Growth, proline accumulation and peroxidase activity in maize seedlings under osmotic stress. Acta Physiologie Plantarum 2013;35 233–239.

[60] Gu J., Yin X., Struik PC., Stomph TJ., Wang H. Using chromosome introgression lines to map quantitative trait loci for photosynthesis parameters in rice (*Oryza sativa* L.) leaves under drought and well-watered field conditions. Journal of Experimental Botany 2012;63(1) 455–469.

[61] Bernier J., Serraj R., Kumar A., Venuprasad R., Impa S., Veeresh Gowda RP., Oane R., Spaner D., Atlin G. The large-effect drought-resistance QTL qtl12.1 increases water uptake in upland rice. Field Crops Research 2009;110 139–146.

[62] Nikolic A., Andjelković V., Dodig D., Mladenović-Drinić S., Kravić N., Ignjatović-Micić D. Identification of QTL-s for drought tolerance in maize, II: Yield and yield components. Genetika 2013;45(2) 341–350.

[63] Kantar M., Lucas SJ., Budak H. miRNA expression patterns of *Triticum dicoccoides* in response to shock drought stress. Planta 2011;233(3) 471–484.

[64] Yeo ET., Kwon HB., Han SE., Lee JT., Ryu JC., Byu MO. Genetic engineering of drought resistant potato plants by introduction of the trehalose-6-phosphate synthase (TPS1) gene from *Saccharomyces cerevisiae*. Molecules and Cells 2000;10(3) 263–268.

[65] Kramer PJ., Boyer JS. Photosynthesis and water availability. In: Water Relations of Plants and Soils. USA, Academic Press, Elsevier Science 1995, p. 315–319.

[66] Liao C-T., Lin C-H. Physiological adaptation of crop plants to flooding stress. Proceedings of the National Science Council, Republic of China (B) 2001;25(3) 148–157.

[67] Ashraf M., Arfan M. Gas exchange characteristics and water relations in two cultivars of *Hibiscus esculentus* under waterlogging. Biologia Plantarum 2005;49(3) 459–462.

[68] Bramley H., Turner NC., Turner DW., Tyerman SD. The contrasting influence of short-term hypoxia on the hydraulic properties of cells and roots of wheat and lupin. Functional Plant Biology 2010;37 183–193.

[69] Bailey-Serres J., Voesenek LA. Life in the balance: a signaling network controlling survival of flooding. Current Opinion in Plant Biology 2010;13(5) 489–494.

[70] Garnczarska M., Bednarski W. Effect of a short-term hypoxic treatment followed by re-aeration on free radicals level and antioxidative enzymes in lupine roots. Plant Physiology and Biochemistry 2004;42(3) 233–240.

[71] Subbaiah CC., Sachs MM. Molecular and cellular adaptations of maize to flooding stress. Annals of Botany 2003;91(2) 119–127.

[72] Mustroph A., Albrecht G. Tolerance of crop plants to oxygen deficiency stress: fermentative activity and photosynthetic capacity of entire seedlings under hypoxia and anoxia. Physiologia Plantarum 2003;117(4) 508–520.

[73] Setter TL., Waters I. Review of prospects for germplasm improvement for waterlogging tolerance in wheat, barley and oats. Plant and Soil 2003;253(1) 1–34.

[74] Dakora FD., Phillips DA. Root exudates as mediators of mineral acquisition in low-nutrient environments. Plant and Soil 2002;245(1) 35–47.

[75] Läuchli A., Grattan SR. Plant growth and development under salinity stress. Jenks MA., Hasegawa PM. Jain SM. (ed.) Advances in Molecular Breeding toward Drought and Salt Tolerant Crops. Dordrecht, Netherlands, Springer 2007, 1–32.

[76] Meloni DA., Oliva MA., Ruiz HA., Martinez CA. Contribution of proline and inorganic solutes to osmotic adjustment in cotton under salt stress. Journal of Plant Nutrition 2001;24(3) 599–612.

[77] Essa TA. Effect of salinity stress on growth and nutrient composition of three soybean (*Glycine max* L. Merrill) cultivars. Journal of Agronomy and Crop Science 2002;188(20) 86–93.

[78] Schabes FI., Sigstad EE. Calorimetric studies of quinoa (*Chenopodium quinoa* Willd.) seed germination under saline stress conditions. Thermochimica Acta 2005;428 71–75.

[79] Kerepesi I., Galiba G. Osmotic and salt stress-induced alteration in soluble carbohydrate content in wheat seedlings. Crop Science 2000;40 482–487.

[80] Zhao FG., Qin P. Protective effect of exogenous polyamines on root tonoplast function against salt stress in barley seedlings. Plant Growth Regulation 2004;42(2) 97–103.

[81] Tanou G., Molassiotis A., Diamantidis G. Hydrogen peroxide- and nitric oxide-induced systemic antioxidant prime-like activity under NaCl-stress and stress-free conditions in citrus plants. Journal of Plant Physiology 2009;166(17) 1904–1913.

[82] Parvaiz A., Satyawati S. Salt stress and phyto-biochemical responses of plants—a review. Plant, Soil and Environment 2008;54(3) 89–99.

[83] Munns R., Tester M. Mechanisms of salinity tolerance. Annual Review of Plant Biology 2008;59 651–681.

[84] Jorge RA., Arrunda P. Aluminium induced organic acids exudation by roots of aluminium tolerant maize. Phytochemistry 1997;45 675–681.

[85] Sumner ME., Noble AD. Soil acidification: the world story. In: Rengel Z. (ed.) Handbook of Soil Acidity. Marcel Dekker, New York 2003, p. 1-28.

[86] Eckhard G., Horst WJ., Neumann E. Adaptation of plants to adverse chemical soil conditions. In: Marschner P. (ed.) Marschner's Mineral Nutrition of Higher Plants (Third Edition) 2012, p. 409–472.

[87] Samac DA., Tesfaye M. Plant improvement for tolerance to aluminum in acid soils—a review. Plant Cell, Tissue and Organ Culture 2003;75(3) 189–207.

[88] Pandurovic Z., Dragicevic V., Glamoclija Dj., Dumanovic Z. Possibilities of maize cropping for feed on acid soil. Journal of Animal and Veterinary Advances 2013;12(7) 813–822.

[89] Pertusatti J., Prado AGS. Buffer capacity of humic acid: thermodynamic approach. Journal of Colloid and Interface Science 2007;314(2) 484–489.

[90] Tesfaye M., Temple SJ., Allan DL., Vance CP., Samac DA. Overexpression of malate dehydrogenase in transgenic alfalfa enhances organic acid synthesis and confers tolerance to aluminum. Plant Physiology 2001;127(4) 1836–1844.

[91] Powles SB., Yu Q. Evolution in action: plants resistant to herbicides. Annual Review of Plant Biology 2010;61 317–347.

[92] Stefanovic L., Simic M., Rosulj M., Vidakovic M., Vancetovic J., Milivojevic M., Misovic M., Selakovic D., Hojka Z. Problems in weed control in serbian maize seed production. Maydica 2007;52 277–280.

[93] Bigot A., Fontaine F., Clément C., Vaillant-Gaveau N. Effect of the herbicide flumioxazin on photosynthetic performance of grapevine (*Vitis vinifera* L.). Chemosphere 2007;67(6) 1243–1251.

[94] Rutherford AW., Krieger-Liszkay A. Herbicide-induced oxidative stress in photosystem II. Trends in Biochemical Sciences 2001;26(11) 648–653.

[95] Castro AJ., Carapito C., Zorn N., Magné C., Leize E., Van Dorsselaer A., Clément C. Proteomic analysis of grapevine (*Vitis vinifera* L.) tissues subjected to herbicide stress. Journal of Experimental Botany 2005;56(421) 2783–2795.

[96] Dragičević V., Simić M, Stefanović L., Sredojević S. Possible toxicity and tolerance patterns towards post-emergence herbicides in maize inbred lines. Fresenius Environmental Bulletin 2010;19(8) 1499–1504.

[97] Dragićević V., Simić M., Sečanski M., Cvijanović G., Nišavić A. Study of the susceptibility of maize lines to some sulfonylurea herbicides. Genetika 2012;44(2) 355–366.

[98] Brankov M., Dragicevic V., Simic M., Spasojevic I. Dynamics of soluble protein content and grain yield in maize inbred lines influenced by foramsulfuron. Conference Proceedings Fifth International Scientific Agricultural Symposium "Agrosym 2014" October 23–26, Jahorina, Republic of Srpska, Bosnia, 2014, p. 497–450.

[99] Carvalho SJP., Nicolai M., Rodrigues Ferreira R., Oliveira Figueira AV., Christoffoleti PJ. Herbicide selectivity by differential metabolism: considerations for reducing crop damages. Sciencia Agricola (Piracicaba, Brazil) 2009;66(1)136–142

[100] Dragičević V., Simić M., Brankov M. Spasojević I., Sečanski M., Kresović B. Thermodynamic characterization of early phytotoxic effects of sulfonylurea herbicides to maize lines. Pesticides & Phytomedicine 2012;27(3) 231–237.

[101] Sun, WQ. Methods for the study of water relations under desiccation stress, In: Black M., Pritchard HW. (ed.), Desiccation and Survival in Plants: Drying Without Dying. New York, USA, CABI Publishing 2002, p. 47–91.

[102] Dragicevic V., Sredojevic S. Thermodynamics of seed and plant growth, In: Piraján JCM. (ed.) Thermodynamics—Systems in Equlibrium and Non-Equilibrium. Rijeka, Croatia, InTech 2011 1–20.

[103] Dragičević V., Simić M., Sredojević S. The influence of herbicides on changes of the phytic and the inorganic phosphorus during starting growth of maize inbred lines. Plant Protection 2010;61(3) 199–206.

[104] Vančetović J., Simić M., Božinović S. The use of CTM (cycloxydim tolerant maize) mutation in maize weeds control. Tomlekova NB., Kozgar MI. Wani MR. (eds) Mutagenesis: Exploring Novel Genes and Pathways. Wageningen Academic Publishers 2014, p. 203–214.

[105] Zhu Y., Yu H., Wang J., Fang W., Yuan J., Yang Z. Heavy metal accumulations of 24 asparagus bean cultivars grown in soil contaminated with Cd alone and with multiple metals (Cd, Pb, and Zn). Journal of Agricultural and Food Chemistry 2007;55(3) 1045–1052.

[106] Babula P., Adam V., Opatrilova R., Zehnalek J., Havel L., Kizek R. Uncommon heavy metals, metalloids and their plant toxicity: a review. Environmental Chemistry Letters 2008;6(4) 189–213.

[107] Hossain MA., Piyatida P., da Silva JAT., Fujita M. Molecular mechanism of heavy metal toxicity and tolerance in plants: central role of glutathione in detoxification of reactive oxygen species and methylglyoxal and in heavy metal chelation. Journal of Botany 2012 Article ID 872875, 37 pages.

[108] Cobbett CS. Phytochelatin biosynthesis and function in heavy-metal detoxification. Current Opinion in Plant Biology 2000;3(3) 211–216.

[109] Clemens S. Toxic metal accumulation, responses to exposure and mechanisms of tolerance in plants. Biochimie 2006;88(11) 1707–1719.

[110] Fässler E., Robinson BH., Stauffer W., Gupta SK., Papritz A., Schulin R. Phytomanagement of metal-contaminated agricultural land using sunflower, maize and tobacco. Agriculture, Ecosystems and Environment 2010;136(1–2) 49–58.

[111] Zhou Y-F., Haynes RJ. Sorption of heavy metals by inorganic and organic components of solid wastes: significance to use of wastes as low-cost adsorbents and immobilizing agents. Critical Reviews in Environmental Science and Technology 2010;40(11) 909–977.

Chapter 9

The Planck Power – A Numerical Coincidence or a Fundamental Number in Cosmology?

Jack Denur

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/61642

Abstract

The Planck system of units has been recognized as the most fundamental such system in physics ever since Dr. Max Planck first derived it in 1899. The Planck system of units in general, and especially the Planck power in particular, suggest a simple and interesting cosmological model. Perhaps this model may at least to some degree represent the real Universe; even if it does not, it seems interesting conceptually. The Planck power equals the Planck energy divided by the Planck time, or equivalently the Planck mass times c^2 divided by the Planck time. We show that the nongravitational mass-energy of our local region (L-region) of the Universe is, at least approximately, to within a numerical factor on the order of 2, equal to the Planck power times the elapsed cosmic time since the Big Bang. This result is shown to be consistent, to within a numerical factor on the order of 2, with results obtained via alternative derivations. We justify employing primarily L-regions within an observer's cosmological *event* horizon, rather than O-regions (observable regions) within an observer's cosmological *particle* horizon. Perhaps this might imply that as nongravitational mass-energy leaves the cosmological event horizon of our L-region via the Hubble flow, it is replaced at the rate of the Planck power and at the expense of negative gravitational energy. Thus the total mass-energy of our L-region, and likewise of all L-regions, is conserved at the value zero. Some questions concerning the Second Law of Thermodynamics and possible thwarting of the heat death of the Universe predicted thereby, whether via Planck-power input or via some other agency, are discussed.

Keywords: Planck system of units, L-regions (local regions), O-regions (observable regions), comoving frame, Second Law of Thermodynamics, heat death, Planck power versus heat death, low-entropy boundary conditions versus heat death, kinetic versus thermodynamic control, kinetic control versus heat death, minimal Boltzmann brains, extraordinary observers.

1. Introduction

In Sect. 2 we define and distinguish between local regions (L-regions) within an observer's cosmological *event* horizon and observable regions (O-regions) within an observer's cosmological *particle* horizon, of the Universe, and justify primarily employing L-regions. In Sect. 3 we discuss the importance of the Planck system of units, which has been recognized as the most fundamental such system in physics ever since Dr. Max Planck first derived it in 1899. We then consider a possibly important role of the Planck system of units, especially of the Planck power, in cosmology. Perhaps the ensuing cosmological model may at least to some degree represent the real Universe; even if it does not, it seems interesting conceptually. The Planck power equals the Planck energy divided by the Planck time, or equivalently the Planck mass times c^2 divided by the Planck time. In Sect. 3 we show that the nongravitational mass-energy of our local region (L-region) of the Universe is, at least approximately, to within a numerical factor on the order of 2, equal to the Planck power times the elapsed cosmic time since the Big Bang. This result is shown to be consistent, to within a numerical factor on the order of 2, with results obtained via alternative derivations. We consider the possible inference that as nongravitational mass-energy leaves the cosmological event horizon of our L-region via the Hubble flow, it is replaced at the rate of the Planck power and at the expense of negative gravitational energy. The problem of consistency with astronomical and astrophysical observations is discussed in Sect. 4. In Sects. 3 and 4 we consider only nonoscillating cosmologies (except for brief parenthetical mentions of oscillating ones in the second-to-last paragraph of Sect. 4). In Sects. 5–8 we consider both nonoscillating and oscillating cosmologies. Some questions concerning the Second Law of Thermodynamics and possible thwarting of the heat death predicted thereby are discussed with respect to the Planck power in Sect. 4, with respect to cosmology in general and minimal Boltzmann brains in particular in Sect. 5, with respect inflation to in Sect. 6, and with respect to kinetic versus thermodynamic control in Sects. 4 and 7. (We discuss possible thwarting of the heat death with respect to kinetic versus thermodynamic control mainly as regards the Planck power in particular in Sect. 4 but more generally in Sect. 7.) A brief review concerning the Multiverse, and some alternative viewpoints, are given in Sect. 8.

2. L-regions and O-regions

In this chapter we will consider primarily *local* regions or L-regions of the Universe rather than *observable* regions or O-regions [1] thereof, although we will also consider O-regions as necessary [1].[1] We now define and distinguish between L-regions and O-regions, and justify primarily employing L-regions, as opposed to O-regions only occasionally, as necessary [1]. Let R be the radial ruler distance or proper distance [2] to the boundary of our L-region, that is to our cosmological *event* horizon [3], where the Hubble flow is at c, the speed of light in vacuum; beyond this horizon it exceeds c. Thus if the Hubble constant $H(\tau)$ does not vary with cosmic time [4,5] τ and is always equal to its present value H_0, then light emitted at the *present* cosmic time [4,5] τ_0 by sources beyond our cosmological event horizon [2,3] and hence beyond our L-region can never reach us. Likewise, light emitted at the *present* cosmic time [4,5] τ_0 by us can never reach them. Also, if the Hubble constant $H(\tau)$ does not vary with cosmic time [4,5] τ and is always equal to its present value H_0, then our cosmological event horizon [2,3] is always at fixed ruler distance $R_0 = c/H_0$ away and hence our L-region

[1] (Re: Entry [1], Ref. [1]) In Ref. [1] observable regions of the Universe are referred to as O-regions for short. We have followed this notation with respect to both O-regions and local regions (L-regions) in this chapter, with L-regions being of primary interest to us.

http://dx.doi.org/10.5772/61642

of the Universe is always of fixed size. [We denote the value of a given quantity Q *today* (at the *present* cosmic time τ_0) by Q_0 and its value at *general* cosmic time τ by $Q(\tau)$.]

Light emitted at *past* cosmic times $\tau < \tau_0$ (but not too far in the past) by sources now beyond (but not too far beyond) our cosmological event horizon [$R_0 = c/H_0$ always if $H(\tau) = H_0$ always] and hence beyond our L-region but still within our O-region [1–3] can reach us, because when this light was emitted these sources were still within our L-region. Likewise, light emitted in the *past* $\tau < \tau_0$ (but not too far in the past) by us can reach them. The boundary of our O-region of the Universe is our cosmological *particle* horizon [1–3]. The boundary of our O-region (our cosmological particle horizon) is further away than the boundary of our L-region (our cosmological particle horizon) [1–3]. If $H(\tau) = H_0$ always, not only is the boundary of our O-region currently at ruler distance $\mathfrak{R}_0 > R_0 = c/H_0$ but $\mathfrak{R}(\tau)$ gets further away with increasing cosmic time τ [4,5], while the boundary of our L-region $R(\tau)$ always remains fixed at $R_0 = c/H_0$. The fixed size of our (or any) L-region given constant $H(\tau) = H_0$ simplifies our discussions. More importantly, all parts of our (or any) L-region are *always* in casual contact, while outer parts of our (or any) O-region beyond the limit of the corresponding L-region *were* but no longer *are* in causal contact. Hence we will primarily employ L-regions rather than O-regions.

Hubble flow exceeding c may seem to violate Special Relativity. But General Relativity — not Special Relativity — is applicable in cosmology [6]. Special Relativity is applicable only within local inertial frames, and any given observer is not — indeed cannot be — in the same local inertial frame as this observer's cosmological event horizon [1–6] (and even less so as this observer's cosmological particle horizon [1–6]). Thus Hubble flow exceeding c does not violate General Relativity [6]. It should also be noted that the Hubble flow is motion with space rather than through space — every object in the Hubble flow is at rest in the comoving frame [7]. An object's motion, if any, relative to the comoving frame [7] is its *peculiar* motion.[2]

At the 27th Texas Symposium on Relativistic Astrophysics [8], values of the Hubble constant today H_0 from the upper 60s to the low 70s $(\text{km}/\text{s})/\text{Mpc}$ were given [8], so $H_0 \approx 70\,(\text{km}/\text{s})/\text{Mpc}$ splits the difference [8]. These values were essentially unchanged from those obtained shortly preceding this Symposium [9,10]. The *Planck* 2015 results [11] state a value of $H_0 = 68\,(\text{km}/\text{s})/\text{Mpc}$ [11], but this *Planck* 2015 work [11] also cites other recent results that range from the low 60s $(\text{km}/\text{s})/\text{Mpc}$ to the low 70s $(\text{km}/\text{s})/\text{Mpc}$. Thus the value $H_0 = 68\,(\text{km}/\text{s})/\text{Mpc}$ [11] not only is the most reliable and most recent one as of this writing, but it also splits the difference of the range of other recent results cited in this *Planck* 2015 work [11]. Hence we take the Hubble constant today to be $H_0 \doteq 68\,(\text{km}/\text{s})/\text{Mpc} \approx 2.2 \times 10^{-18}\,(\text{km}/\text{s})/\text{km} = 2.2 \times 10^{-18}\,\text{s}^{-1}$ [11].[3]

[2] (Re: Entry [7], Ref. [2]) An observer in the comoving frame (ideally in intergalactic space as far removed as possible from local gravitational fields such as those of galaxies, stars, etc.) sees the 2.7 K cosmic background radiation as isotropic (apart from fluctuations of fractional magnitude $\mathcal{F} \approx 10^{-5}$, which can be "smoothed out" via, say, computer processing to yield a uniform background). But even Earth is a fairly good approximation to the comoving frame: Earth's peculiar motion $\approx 380\,\text{km}/\text{s} \ll c$ (see p. 352 of Ref. [2]) with respect to the cosmic background radiation is fairly slow, and local gravitational fields are fairly weak ($v_{\text{escape}} \ll c$).

[3] (Re: Entries [8]–[11], Refs. [2], [8], [10], and [11]) As per Entries [8]–[11], results for the Hubble constant have improved with time, asymptotically converging onto those provided by Ref. [11]. The results for the Hubble constant as per Ref. [8] are in essential agreement with Entry [9]. The history of values of the Hubble constant also is briefly discussed in Entry [9] and Ref. [10]. Reference [10] surveys the history of values of the Hubble constant determined via work done through 2012. Reference [10] was for sale at the 27th Texas Symposium on Relativistic Astrophysics, held at the Fairmont Hotel in Dallas, Texas, December 8–13, 2013.

3. The Planck power in cosmology

The Planck system of units has been recognized as the most fundamental such system in physics ever since Dr. Max Planck first derived it in 1899 [12–15]. It is based on Planck's reduced constant $\hbar \equiv h/2\pi$ (or Planck's original constant h), the speed of light in vacuum c, and the universal gravitational constant G, with Boltzmann's constant k usually also included. These four fundamental physical constants are seen by *everything*, corresponding to the Planck system of units encompassing *universal* domain. By contrast, for example, the fundamental electric charge is seen only by electrically-charged particles.[4]

The Planck system of units in general, and especially the Planck power in particular, suggest a simple and interesting cosmological model. Perhaps this model may at least to some degree represent the real Universe; even if it does not, it seems interesting conceptually.

Multiply the Planck mass $m_{\text{Planck}} = (\hbar c/G)^{1/2}$ by c^2 to obtain the Planck energy $E_{\text{Planck}} = \left(\hbar c^5/G\right)^{1/2}$ [12–15]. Divide the Planck energy by the Planck time $t_{\text{Planck}} = \left(\hbar G/c^5\right)^{1/2}$ to obtain the Planck power $P_{\text{Planck}} = c^5/G \doteq 3.64 \times 10^{52}\,\text{W} \iff P_{\text{Planck}}/c^2 = c^3/G \doteq 4.05 \times 10^{35}\,\text{kg / s}$ [12–15]. [The dot-equal sign ($\doteq$) means "very nearly equal to."] Note that — unlike the Planck length, mass, energy, time, and temperature $T_{\text{Planck}} = E_{\text{Planck}}/k = \left(\hbar c^5/G\right)^{1/2}/k$, indeed unlike most if not all other Planck units (at least most if not all other useful ones except the Planck speed $l_{\text{Planck}}/t_{\text{Planck}} = c$) — the Planck power (whether or not divided by c^2) does *not* contain $\hbar$, but only G and c. Thus — unlike the Planck length l_{Planck}, Planck mass, Planck energy, Planck time, and Planck temperature, indeed unlike most if not all other Planck units (at least most if not all other useful ones except the Planck speed $l_{\text{Planck}}/t_{\text{Planck}} = c$) — is the Planck power a *classical* quantity *in*dependent of quantum effects, if not absolutely then at least via opposing quantum effects canceling out, as $\hbar$ cancels out in the division $P_{\text{Planck}} = E_{\text{Planck}}/t_{\text{Planck}}$? With respect to the Planck speed $l_{\text{Planck}}/t_{\text{Planck}} = c$ note that c is the fundamental speed in the classical (nonquantum) theories of Special and General Relativity.

Now multiply P_{Planck}/c^2 by the age of the Universe, the elapsed cosmic time [4,5] since the Big Bang, $\tau_0 \approx 4.5 \times 10^{17}\,\text{s} \approx 1.4 \times 10^{10}\,\text{y}$ [11]. This yields an estimate of

$$M_0 \approx \frac{P_{\text{Planck}}\,\tau_0}{c^2} \approx 1.8 \times 10^{53}\,\text{kg} \tag{1}$$

for the mass of our L-region (not considering the negative gravitational energy). But $M_0 \approx 1.8 \times 10^{53}\,\text{kg}$ is of order-of-magnitude agreement with an estimate of M_0 assuming that the mass-energy density of of our L-region of the Universe [1,3] equals the critical density ρ_{crit} [16], as seems to be the case if not exactly then at least to within a very close approximation. The density critical density ρ_{crit} corresponds to the borderline between ever-expanding and oscillating Universes given vanishing cosmological constant, i.e., $\Lambda = 0$,

[4] (Re: Entries [12] and [15], Refs. [12] and [15]) A concise listing of Planck units and other useful data, entitled "Some Useful Numbers in Conventional and Geometrized Units," is provided in the back endcover of Ref. [12]. In this back endcover of Ref. [12] the Planck length is referred to as the Planck distance (elsewhere in Ref. [12] it is referred to as the Planck length) and the Planck power is referred to as the emission factor. Reference [12] cites Ref. [13] as the most important work in the derivation of the Planck system of units. Reference [15], like Ref. [12], cites Ref. [13]. Additionally, in Sect. 31.1, Ref. [15] gives a brief historical survey of works deriving the Planck system of units. Reference [15] extends the Planck system of units to also include Boltzmann's constant k.

http://dx.doi.org/10.5772/61642

and to spacetime being flat, and hence space Euclidean, on the largest scales, i.e., to the spatial curvature index being 0 rather than $+1$ or -1, given *any* value of Λ [16–20].[5] The critical density is

$$\rho_{\rm crit} = \frac{3H_0^2}{8\pi G} \approx 8.65 \times 10^{-27} \frac{\rm kg}{\rm m^3}. \tag{2}$$

Applying the most recent and best result for H_0, namely $H_0 \approx 68\,({\rm km\,/\,s})\,/\,{\rm Mpc} \approx 2.2 \times 10^{-18}\,({\rm km\,/\,s})\,/\,{\rm km} = 2.2 \times 10^{-18}\,{\rm s}^{-1}$ [11], yields as an estimate of M_0

$$M_0 \approx \frac{4\pi}{3}\rho_{\rm crit}R_0^3 = \frac{4\pi}{3}\frac{3H_0^2}{8\pi G}\left(\frac{c}{H_0}\right)^3 = \frac{c^3}{2GH_0} \approx 9.2 \times 10^{52}\,{\rm kg}\,. \tag{3}$$

In Eq. (3) we assume that the volume of our L-region is given by the Euclidean value $4\pi R_0^3/3$. But since astronomical observations indicate that spacetime is flat, and hence space is Euclidean, on the largest scales, i.e., that the spatial curvature index is 0 rather than $+1$ or -1, this assumption seems justified [11,16–20]. Is the order-of-magnitude agreement between Eqs. (1) and (3) merely a numerical coincidence? Or does it suggest that the Planck power plays a fundamental role in cosmology — entailing a link between the smallest (Planck-length and Planck-time) and largest (cosmological) scales?

While there is order–of magnitude agreement between Eqs. (1) and (3), there is a discrepancy between them by a factor of ≈ 2. That is, Planck-power input as per Eq. (1) seems to imply $\rho \approx 2\rho_{\rm crit}$. Since in this era of precision cosmology all quantities in Eqs. (1)–(3) are known far more accurately than to within a factor of 2, it seems that this factor of ≈ 2 can*not* simply be dismissed. But we admit that we have no explanation for this factor of ≈ 2. Furthermore, we will see that Eqs. (5)–(7) seem to imply a discrepancy with Eq. (1) by a factor of $\approx 3/2$ in the opposite direction, i.e., that Planck-power input as per Eq. (1) seems to imply $\rho \approx 2\rho_{\rm crit}/3$. Such discrepancies by numerical factors on the order of 2 may prove our Planck-power hypothesis to be wrong. At the very least they prove that even if it is right *in general* it is only an *introductory* hypothesis whose *details* still need to be understood. Then again, perhaps because there *is* consistency to within a small numerical factors of $O \sim 2$, our Planck-power hypothesis may be correct *in general* as an *introductory* hypothesis, even though, even if correct *in general*, its *details* still need to be understood.

Do our considerations so far in this Sect. 3 suggest that, even though the Universe certainly began with the Big Bang, there has been since the Big Bang mass-energy input, at least on the average, at the Planck power, into our L-region of the Universe? We list several alternative proposals for such input (this list probably is not exhaustive): (a) steady-state-theory mass-energy input *ex nihilo* [21–23], (b) mass-energy input *ex nihilo* via other means [24,25], (c) mass-energy input at the expense of negative gravitational energy [26–32] rather than *ex nihilo*, or (d) mass-energy input at the expense of nongravitational negative energy, for example, at the expense of the negative-energy C field in some versions of the steady-state theory [33–35]. If at the expense of negative gravitational energy as per proposal (c), then *forever* the total (mass plus gravitational) energy of our L-region, and likewise of any L-region, of the Universe, and hence of the Universe as a whole, is conserved at the value

[5] (Re: Entry [20], Ref. [14]) The critical density and density parameter are employed on various occasions throughout Chap. 29 on cosmology in Ref. [14].

zero [26–32]. (There are "certain 'positivity' theorems ... which tell us that the total energy of a system, including the 'negative gravitational potential energy contributions' ..., cannot be negative [32]." But positivity theorems do seem to allow the total energy of a system, including the negative gravitational energy, to be strictly zero. Also, perhaps positivity theorems need necessarily apply only for isolated sources in asymptotically-flat spacetime.) In this chapter we will mainly presume proposal (c) from the immediately preceding list, for the following reasons: (i) Unlike proposals (a) and (b), proposal (c) entails no violation of the First Law of Thermodynamics (conservation of mass-energy). (ii) Negative gravitational energy is *known* to exist, unlike the negative-energy C field of proposal (d), which was perhaps introduced at least partially *ad hoc* to render the steady-state theory consistent with the First Law of Thermodynamics (conservation of mass-energy). Moreover, unlike gravity, the C field not only has never been observed, but also entails difficulties of its own [34,35]. (iii) We will show that proposal (c) need not be inconsistent with the observed features of the Universe.

The Universe clearly shows evolutionary rather than steady-state [21–23,33–35] behavior since the Big Bang. But it could stabilize to a steady state in the future. It could already now be thus stabilizing or even thus stabilized in the very recent past with as yet no or at most very limited observational evidence that might be suggestive of such stabilization. Thus even if there is steady-state-type creation of mass-energy since the Big Bang at the rate of the Planck power (we presume, in light of the immediately preceding paragraph, most likely at the expense of the Universe's negative gravitational energy), perhaps this might be compatible with the observed evolutionary behavior of the Universe since the Big Bang. (This point and related ones will be discussed in more detail in Sect. 4.)

Although General Relativity is required for an accurate consideration of the Universe's gravity, the following Newtonian approximation may be valid as an order-of-magnitude estimate [26–32]. Such an estimate is suggestive in favor of Planck-power input at the expense of negative gravitational energy [26–32], which does not require a violation of the First Law of Thermodynamics (conservation of mass-energy) [26–32], as opposed to Planck-power input *ex nihilo* [21–25], which would require such a violation, or via C-field input, the C field never having been observed and also entailing its own difficulties [34,35]. In accordance with the last paragraph of Sect. 2, we take the Hubble constant today to be $H_0 \doteq 68\,(\mathrm{km/s})/\mathrm{Mpc} \approx 2.2 \times 10^{-18}\,(\mathrm{km/s})/\mathrm{km} = 2.2 \times 10^{-18}\,\mathrm{s}^{-1}$ [8–11]. Thus neglecting any variation of $H(\tau)$ with τ, $\tau_0 = 1/H_0 \approx 4.5 \times 10^{17}\,\mathrm{s}$ consistently with the previously given value, and the ruler radius of our L-region of the Universe is $R_0 = c\tau_0 = c/H_0 \approx 1.4 \times 10^{23}\,\mathrm{km} = 1.4 \times 10^{26}\,\mathrm{m}$. The positive mass-energy of our L-region of the Universe within our cosmological event horizon is M_0c^2 and the negative Newtonian gravitational energy of our L-region is $\approx -GM_0^2/R_0$. Hence in the Newtonian approximation setting the total energy equal to zero yields [26–32]

$$\begin{aligned} E_{\mathrm{total}} = E_{\mathrm{mass}} + E_{\mathrm{gavitational}} &= 0 \\ \implies M_0c^2 - \frac{GM_0^2}{R_0} &= 0 \\ \implies \frac{M_0}{R_0} &= \frac{c^2}{G}. \end{aligned} \tag{4}$$

http://dx.doi.org/10.5772/61642

Applying our previously derived values of M_0 and R_0 yields $M_0/R_0 \approx 1.8 \times 10^{53}\,\text{kg}\,/1.36 \times 10^{26}\,\text{m} \approx 1.32 \times 10^{27}\,\text{kg}\,/\,\text{m}$. We have $c^2/G \doteq 1.35 \times 10^{27}\,\text{kg}\,/\,\text{m}$. Thus Eq. (4) is fulfilled as closely as we can expect, especially given that our Newtonian approximation should be expected to provide only order-of-magnitude estimates, and also perhaps because (even after an initial fast inflationary stage) $H(\tau)$ may not be strictly constant.

There is yet another order-of-magnitude result that is consistent with our Planck-power hypothesis. Applying Eq. (1), rate of Planck-power mass input into our L-region is

$$\left(\frac{dM}{d\tau}\right)_{\text{in}} \approx \frac{M_0}{\tau_0} \approx \frac{P_{\text{Planck}}}{c^2} = \frac{\frac{c^5}{G}}{c^2} = \frac{c^3}{G}. \tag{5}$$

Letting ρ be the average density of our L-region, the rate of Hubble-flow mass-export from our L-region is

$$\left(\frac{dM}{d\tau}\right)_{\text{out}} = 4\pi R_0^2 \rho c = 4\pi \left(\frac{c}{H_0}\right)^2 \rho c = \frac{4\pi\rho c^3}{H_0^2}. \tag{6}$$

In Eq. (6) we assume that the surface area bounding our L-region is given by the Euclidean value $4\pi R_0^2$. But since astronomical observations indicate that spacetime is flat, and hence space is Euclidean, on the largest scales, i.e., that the spatial curvature index is 0 rather than $+1$ or -1, this assumption seems justified [11,16–20]. For steady-state to obtain we must have

$$\begin{aligned}
&\left(\frac{dM}{d\tau}\right)_{\text{net}} = \left(\frac{dM}{d\tau}\right)_{\text{in}} - \left(\frac{dM}{d\tau}\right)_{\text{out}} = 0 \\
&\implies \frac{c^3}{G} - 4\pi R_0^2 \rho c = \frac{c^3}{G} - \frac{4\pi\rho c^3}{H_0^2} = 0 \\
&\implies \frac{1}{G} - \frac{4\pi\rho}{H_0^2} = 0 \\
&\implies \rho = \frac{c^2}{4\pi G R_0^2} = \frac{H_0^2}{4\pi G} \approx 5.8 \times 10^{-27}\,\frac{\text{kg}}{\text{m}^3}.
\end{aligned} \tag{7}$$

The numerical value for ρ obtained in the last line of Eq. (7) is in order-of-magnitude agreement with ρ_{crit} as per Eq. (2), as well as in order-of-magnitude agreement with observations.

Recalling the second paragraph following that containing Eqs. (1)–(3), note that Eqs. (5)–(7) seem to imply a discrepancy with Eq. (1) by a factor of $\approx 3/2$ in the opposite direction from the discrepancy with Eq. (1) by a factor of ≈ 2 implied by Eq. (3). Planck-power input as per Eq. (1) seems to imply $\rho \approx 2\rho_{\text{crit}}$, while Eqs. (5)–(7) seem to imply $\rho \approx 2\rho_{\text{crit}}/3$. Since in this era of precision cosmology all quantities in Eqs. (1)–(3) and (5)–(7) are known far more accurately than to within a factor of 2, such discrepancies by numerical factors of $O \sim 2$ may prove our Planck-power hypothesis to be wrong. At the very least they prove that even if it is

right *in general* it is only an *introductory* hypothesis whose *details* still need to be understood. Then again, perhaps because there *is* consistency to within small numerical factors of $O \sim 2$, our Planck-power hypothesis may be correct *in general* as an *introductory* hypothesis, even though, even if correct *in general*, its *details* still need to be understood.

While, even accepting discrepancies by a factor of $O \sim 2$, the fulfillment of Eqs. (1)–(7) does not constitute proof of Planck-power input, it at least seems suggestive. Could Planck-power input, if it exists, be a classical process *in*dependent of quantum effects, if not absolutely then at least via opposing quantum effects canceling out, as $\hbar$ cancels out in the division $P_{\text{Planck}} = E_{\text{Planck}}/t_{\text{Planck}}$? Note that perhaps similar canceling out obtains with respect to the Planck speed $l_{\text{Planck}}/t_{\text{Planck}} = c$: c is the fundamental speed in the classical (nonquantum) theories of Special and General Relativity.

According to most current cosmological models, there probably have been one-time initial mass-energy inputs, for example associated with phase transitions ending fast-inflationary stages during the very early history of the Universe [36–40]. We should note that while the majority opinion is certainly in favor of inflation [36–50], there is some dissent [36–50]. (The difficulty in squaring inflation with the Second Law of Thermodynamics, and a possible resolution of this difficulty, will be discussed in Sect. 6.) Observational evidence that in early 2014 initially seemed convincing for inflation in general [41,42], albeit possibly ruling out a few specific types of inflation [41,42,47], has been questioned [43–50], but not disproved [43–50].[6] Moreover, even if an inflationary model is correct, recent observational findings disfavor simple models of inflation, such as quadratic and natural inflation [47]. Even if such one-time initial mass-energy inputs [36–40] occurred, could sustained mass-energy input then continue indefinitely such that the Planck power is at least a *floor* below which the *average* rate of mass-energy input into our L-region of the Universe cannot fall? It at least *appears* not to have fallen below this floor [16]. By the cosmological principle [51], if this is true of our L-region of the Universe then it must be true of *any* L-region thereof.

Thus our Planck-power hypothesis at least *appears* to entail a link between the smallest (Planck mass and Planck time) and largest (cosmological) scales, rather than being merely a numerical coincidence.

4. Planck power and kinetic control versus heat death: Big-Bang-initiated evolution merging into steady state?

In the simplest ever-expanding cosmologies, the Universe begins with a Big Bang and expands forever, with flat geometry on the largest scales, and with the Hubble constant $H(\tau)$ not varying with cosmic time [4,5] τ and always equal to its present value H_0. As the Universe expands the Hubble flow carries mass-energy past the cosmological event horizon of our L-region of the Universe. But this "loss" is replaced by new positive mass-energy continually created within our L-region of the Universe *forever* at the rate of the Planck power and at the expense of our L-region's negative gravitational energy. This compensates

[6] (Re: Entry [48], Ref. [48]) Reference [48] shows that previous measurements of the acceleration of the Universe's expansion may require reconsideration, owing to discrepancies between visible-light and UV observations of type 1a supernovae.

http://dx.doi.org/10.5772/61642

for "losses" streaming past the cosmological event horizon of our L-region of the Universe via the Hubble flow — and does so consistently with the First Law of Thermodynamics (conservation of mass-energy) [26–32]. Because by the cosmological principle [51] our L-region is nothing special, the same is true of *any* L-region [3] of the Universe. Thus *forever* the total (mass plus gravitational) energy of our L-region, and likewise of any L-region, of the Universe, and hence of the Universe as a whole, is conserved at the value zero [26–32].

There is needed a mechanism whereby a sufficiently large fraction f of the Planck-power mass-energy input within the cosmological event horizon of our L-region, and of any L-region, of the Universe is produced in the form of hydrogen [52–57] (as in the original steady-state theory [21–23,33–35]), so that there will be fuel for stars [52–57]. Then there will *always* be stars [52–57], planets, and life — not only at the periphery but *even at the center* [58] of our island Universe [1] and likewise of every other island Universe [1] in the Multiverse [52–58]. Then the heat death predicted by the Second Law of Thermodynamics [59–63] will be thwarted not only at the periphery but *even at the center* [58] of our island Universe and likewise of every other island Universe in the Multiverse.[7] The required sufficiently large fraction f is actually quite small. An order-of-magnitude estimate of the total number of stars within the cosmological event horizon [3] of our L-region of the Universe is $\sim 10^{22}$ [64]. By the cosmological principle [51] our L-region of the Universe is nothing special, so there is no reason to suspect a substantially different total number of stars in other L-regions. The Sun's luminosity is $L_{\text{Sun}} = 3.828 \times 10^{26}\,\text{W} \approx 10^{-26} P_{\text{Planck}}$ [56,57]. Thus if the average star were as luminous as the Sun then the total luminosity of $\sim 10^{22}$ stars would be $L_{\text{total}} \sim 10^{-26} P_{\text{Planck}} \times 10^{22} \sim 10^{-4} P_{\text{Planck}}$, implying that we require $f \sim L_{\text{total}}/P_{\text{Planck}} \sim 10^{-4}$ [56,57,64]. But the average star is considerably less luminous than the Sun [56,57], so the best order-of-magnitude estimate is perhaps $L_{\text{total}} \sim 10^{-5} P_{\text{Planck}}$, implying that we require only $f \sim 10^{-5}$ [56,57,64]. (Properties of the Sun, including its luminosity, are given in both conventional and geometrized units in the inside back cover of Ref. [12], and in conventional units in Appendix A in the inside front cover of Ref. [14] and in Table 8.1 on p. 219 of Ref. [10].) This small value $f \sim 10^{-5}$ is sufficient to sustain star formation *forever* not only at the periphery but *even at the center* [58] of our island Universe [1] and likewise of every other island Universe [1] in the Multiverse [52–58]. The remainder of the Planck-power input would be in forms other than hydrogen (perhaps traces of heavier elements, elementary particles of normal and/or dark matter, dark energy, etc.?).

But perhaps the *simplest* mode of Planck-power input is *initially* in the form of the *simplest* possible type of dark energy, corresponding to *positive constant* Λ — a *positive* cosmological *constant*. *Constancy* of Λ is required for *constancy* of Planck-power input initially in the form of Λ. *Positivity* of Λ seems to be required for *positivity* of Planck-power input being initially in the form of Λ, because negative Λ corresponds to contraction of space and hence to diminution of Λ-mass-energy. Thus the *simplest* possible type of dark energy, corresponding to *positive constant* Λ — a *positive* cosmological *constant* — is perhaps the type of dark energy that is most easily reconcilable with Planck-power input, in particular with

[7] (Re: Entry [63], Ref. [63]) Reference [63] considers various aspects of the Second Law of Thermodynamics and its relation to the arrow of time and to cosmology. Reference [63] was for sale at the 27th Texas Symposium on Relativistic Astrophysics, held at the Fairmont Hotel in Dallas, Texas, December 8–13, 2013.

constancy of Planck-power input. Moreover, *constant* Λ — a cosmological *constant* — is the *only, unique,* choice for Λ that can be put on the left-hand (geometry) side of Einstein's field equations without altering their symmetric and divergence-free form [65–67], "belonging to the field equations much as an additive constant belongs to an indefinite integral [65–67].[8] Nevertheless the current trend is to put Λ on the right-hand (mass-energy-stress) side of Einstein's field equations, which allows more freedom [66]. But if Λ is put on the right-hand side "the rationale for its uniqueness then disappears: it no longer needs to be a divergence-free 'geometric' tensor, built solely from the $g_{\mu\nu}$ ... the geometric view of Λ ... is undoubtedly the simplest [66]". Thus we might speculate about a link between constancy of Λ as a (positive) cosmological constant [65–67] and constancy of (positive) Planck-power input: Perhaps Planck-power input occurs *initially* as (positive) cosmological-constant Λ, with $f \sim 10^{-5}$ thereof, then hopefully, somehow, via an as-yet-unknown mechanism, being transformed into hydrogen. It is important to note that — unlike equilibrium blackbody radiation — (positive) cosmological-constant-Λ dark energy seems to be at less than, indeed at far less than, maximum entropy. Thus there seems to be more than enough entropic "room" for $f \sim 10^{-5}$ of positive-cosmological-constant-Λ dark energy to decay into hydrogen, without requiring decay all the way to iron. Positive-cosmological-constant-Λ Planck-power input thus seems to offer the benefits but not the liabilities of the steady-state theory [21–23,33–35] [violation of mass-energy conservation without the C field (which has never been observed) and which also entails other difficulties with it [34,35] — recall the paragraph immediately following that containing Eq. (1)]. Positive cosmological-*constant* Λ also implies, or at least is consistent with, *constant* $H(\tau) = H_0$ at all cosmic times τ, and hence a fixed size of our L-region, with its boundary (event horizon [2,3]) $R(\tau)$ always fixed at $R_0 = c/H_0$. Thus to sum up this paragraph, the *simplest* model *overall* seems to entail (a) positive-cosmological-constant Λ, (b) Planck-power input *initially* as positive-cosmological-constant Λ at the expense of negative gravitational energy, with (c) $f \sim 10^{-5}$ of Planck-power input, then hopefully, somehow, via an as-yet-unknown mechanism, being transformed into hydrogen. We note that the most reliable and most recent astronomical and astrophysical observations and measurements as of this writing are consistent with *positive* cosmological-*constant*-Λ dark energy [68,69], indeed possibly or even probably *more* consistent with *positive* cosmological-*constant*-Λ dark energy than with any other alternative [68,69]. But, of course, this issue is far from being definitely decided [68,69]. Even though our main point in this chapter most naturally based on positive constant Λ, in Sects. 5–7 some other possibilities for Λ will be qualitatively considered.

We cannot help but notice that temperature fluctuations in the cosmic background radiation have a typical fractional magnitude of $\mathcal{F} \approx 10^{-5}$ [70,71]. The *observed and measured* value $\mathcal{F} \approx 10^{-5}$ [70,71] is obviously far more certain than the *speculated* value $f \sim 10^{-5}$; hence the distinction between the $\approx$ symbol as opposed to the $\sim$ symbol. Although it is unlikely that there is a connection between $\mathcal{F} \approx 10^{-5}$ [70,71] and $f \sim 10^{-5}$, it doesn't seem to hurt if we at least mention this numerical concurrence — just in case there might be a connection.

But the following question arises: Even if there is Planck-power input, why is not *all* of it in a thermodynamically-most-probable maximum-entropy form such as (iron + equilibrium

[8] (Re: Entry [67], Ref. [67]) Reference [67] is cited in the passage from Ref. [2] that we cite in Entry [65].

http://dx.doi.org/10.5772/61642

blackbody radiation) and *none* of it as hydrogen — why is not $f = 0$ [52–57]? If this were the case then the heat death predicted by the Second Law of Thermodynamics [59–63] would *not* be thwarted even *with* Planck-power input. While we are not sure of an answer to this question, we can venture what *prima facie* at least seems to be a reasonable guess: (a) Planck-power input (if it exists) generates equal nonzero quantities of both positive mass-energy and negative gravitational energy starting from (zero positive energy + zero negative energy = zero total energy), and the entropy of (zero positive energy + zero negative energy = zero total energy) is *perforce* zero. (b) Planck-power input is a steady-state but nonequilibrium process that does not allow enough time for complete thermalization of the input from the initial value of zero entropy of (zero positive energy + zero negative energy = zero total energy) to the maximum possible positive entropy of (nonzero positive energy + nonzero negative energy = zero total energy) in a form such as (iron + equilibrium blackbody radiation). That is, Planck-power input is *kinetically rather than thermodynamically controlled* [72–77].[9]

Thus even though, *thermodynamically*, Planck-power input should be in a maximum-entropy form such as (iron + equilibrium blackbody radiation), *kinetically* the reaction

$$\begin{aligned} &\text{zero positive energy} \;+\; \text{zero negative energy} \;=\; \text{zero total energy} \\ &\longrightarrow \text{nonzero positive energy} \;+\; \text{nonzero negative energy} \;=\; \text{zero total energy}. \end{aligned} \tag{8}$$

occurs too quickly to allow thermodynamic equilibrium = maximum entropy to be attained. Yet even Planck-power input initially as positive-cosmological-constant Λ, with a fraction $f \sim 10^{-5}$ of Planck-power input hopefully, somehow, via an as-yet-unknown mechanism, being transformed into hydrogen, entails *some* entropy increase. The entropy increase ΔS that it *does* entail is sufficient to render the probability of its reversal as per Boltzmann's relation between entropy and probability, $\text{Prob}\,(\Delta S) = \exp\,(-\Delta S/k)$, equal to zero for all practical purposes. Thus we are justified in placing only a forward arrow (no reverse arrow) at the beginning of the second line of Eq. (8). Thus Planck-power input entails enough entropy increase to stabilize it and prevent its reversal. But it occurs quickly enough to allow *kinetic control* [72–77] to prevent it from entailing *maximal* entropy increase.

To recapitulate out considerations thus far in Sect. 4: Perhaps the simplest possible Planck-power input is *initially* as positive-cosmological-constant Λ. *Positivity* of Λ is required for *positivity* of Planck-power input, and *constancy* of Λ is required for *constancy* of Planck-power input. *Constancy* of Λ is requisite for Λ to be most simply encompassed within Einstein's field equations [65–67], besides correlating with constancy of Planck-power input. Positive-cosmological-*constant* Λ also implies, or at least is consistent with, *constant* $H\,(\tau) = H_0$ at all cosmic times τ, and hence a fixed size of our L-region, with its boundary

[9] (Re: Entries [72–77], Refs. [72–77]) Kinetic versus thermodynamic control is specifically discussed on pp. 41–42 of Ref. [72], p. 43 of Ref. [73], p. 35 of Ref. [74], pp. 311–312 of Ref. [76], and p. 438 of Ref. [77]. Kinetic versus thermodynamic control are contrasted in Sects. 2.18 and 19.11 of Ref. [75]. Helpful auxiliary material is provided in Sects. 2.15–2.17, 2.19–2.21, and 19.16 of Ref. [75]. Reference [73] does not render Ref. [72] obsolete, because Ref. [72] discusses aspects not discussed in Ref. [73], and vice versa. Likewise, Reference [77] does not render Ref. [76] obsolete, because Ref. [76] discusses aspects not discussed in Ref. [77], and vice versa.

(event horizon [2,3]) $R(\tau)$ always fixed at $R_0 = c/H_0$. But we wish for a fraction $f \sim 10^{-5}$ of Planck-power input hopefully, somehow, via an as-yet-unknown mechanism, being transformed into *hydrogen*. *Hydrogen*, so that stars can have fuel. But why hydrogen? Why not a thermodynamically dead form such as (iron + equilibrium blackbody radiation)? Because kinetically, it would be much more difficult for positive-cosmological-constant Λ to be transformed into a complex atom such as iron than into the simplest one — hydrogen. Thus while *thermodynamic* control would favor iron, if *kinetic* control wins then hydrogen is favored [72–77]. Note that kinetic control is vital not only in initial creation of hydrogen, but also in then preserving hydrogen long enough for it to be of use. It is owing to *kinetic* control that the Sun and all other main-sequence stars fuse hydrogen only to helium, not to iron, and are restrained to doing so slowly enough to give them usefully-long lifetimes. Main-sequence fusion of hydrogen to iron is thermodynamically favored, but kinetically its rate of occurrence is for all practical purposes zero. Thus kinetic control wins, limiting main-sequence fusion to helium and at a slow enough rate to give stars usefully-long lifetimes [72–77]. Indeed it is owing to *kinetic* control that not only hydrogen, but also all other elements except iron, do not instantaneously decay to iron. Kinetic control may also argue against positive-cosmological-constant Λ being *completely* transformed into equilibrium blackbody radiation (without iron). A single hydrogen atom can be created at rest with respect to the comoving frame [7]. By contrast, to conserve momentum, at least two photons must be created simultaneously, which may impose a bottleneck that diminishes the rate of such a process kinetically. Hence $f \sim 10^{-5}$ of the Planck-power input in the positive-but-much-less-than-maximal-entropy form of hydrogen may at least *prima facie* seem plausible. Again it doesn't seem to hurt to at least mention the numerical concurrence between $\mathcal{F} \approx 10^{-5}$ [70,71] and $f \sim 10^{-5}$, even if any connection is unlikely.

Note that a zero value for the initial entropy for would also obtain if Planck-power input were *ex nihilo* [21–25] or at the expense of a negative-energy C field (despite its never having been observed and its other difficulties [34,35]) or other negative-energy field rather than at the expense of negative gravitational potential energy: the entropy of (zero positive energy + zero negative energy = zero total energy) would still *perforce* be zero. Thus our considerations of this Sect. 4, including that of dominance of kinetic over thermodynamic control [72–77], would still be applicable.

Our L-region and O-region clearly manifest evolutionary behavior, for example increasing metallicity [52–55] and a decreasing rate of star formation [52–55]. But our Planck-power hypothesis seems to suggest that its evolutionary behavior could gradually merge towards steady-state behavior. Early in the history of our L-region and O-region, star formation occurred at a much faster rate than now, and stars were on the average much more massive and hence *very* much faster-burning [hydrogen-burning rate $\sim$ (mass of star)^3]. Thus hydrogen was consumed faster than a conversion of $f \sim 10^{-5}$ of Planck-power input could replace it: Stars were burning capital in addition to (Planck-power) income — indeed more capital than income. But with decreasing rate of star formation and decreasing average stellar mass, perhaps a steady-state balance between hydrogen consumption and its replacement via $f \sim 10^{-5}$ of Planck-power input could be approached, with stars living solely on (Planck-power) income. Merging of evolutionary towards steady-state behavior could already be beginning or could have even begun in the very recent past with as yet no or at most very limited observational evidence that might be suggestive of it. If such merging exists then both metallicity and star formation rate could stabilize in the future.

http://dx.doi.org/10.5772/61642

Perhaps they even could already now be stabilizing or have even already begun stabilizing in the very recent past with as yet no or at most very limited observational evidence that might be suggestive of such stabilization in particular, or of such merging in general. This stabilization, if it exists, would require only a small fraction $f \sim 10^{-5}$ of Planck-power as hydrogen to maintain the current status quo in our L-region and O-region. This would allow star formation to continue forever not merely at the peripheries of island Universes, but *even in their central regions* [58]. We note that there is observational evidence that might at least be suggestive of "unexplained" hydrogen [78], which perhaps might qualify as such very limited suggestive observational evidence of merging towards steady-state behavior [78].

It should perhaps be re-emphasized that even Planck-power input as hydrogen entails *some* entropy increase and therefore is thermodynamically irreversible, consistently with the Second Law of Thermodynamics while still thwarting the heat death. The heat death is thus thwarted via *dilution* of entropy as an island Universe [1] expands indefinitely, which is consistent with the Second Law [59–63] — not via *destruction* of entropy, which is not: Planck-power input as hydrogen represents input at *positive but far less than maximum* entropy. Thus Planck-power input (if it exists) defeats the heat death predicted by the Second Law of Thermodynamics [59–63] even though it does not defeat the Second Law itself.

Thus with only $f \sim 10^{-5}$ of the Planck-power input as hydrogen, the heat death predicted by the Second Law of Thermodynamics [59–63] of our L-region of our island Universe [1], and likewise of any L-region of any island Universe [1], is thwarted *forever*. The heat death is thwarted *forever* not only at the periphery but *even at the center* [58] of our and every other island Universe [1]. The heat death is thwarted consistently with, not in violation of, the Second Law of Thermodynamics [59–63]. Hubble flow export of entropy (along with mass-energy) out of our L-region of our island Universe [1], and likewise out of any L-region of any island Universe [1], as its expansion creates more volume *forever*, is compensated *forever* by creation of thermodynamically fresh but still positive-entropy mass-energy — most importantly, hopefully, the fraction $f \sim 10^{-5}$ thereof as hydrogen — via Planck-power input.

Steady-state balance between Planck-power input and Hubble-flow expansion of space can allow both the entropy density and the nongravitational mass-energy density in our L-region of our island Universe [1], and likewise in any L-region of any island Universe [1], to remain constant, even as the total entropy and nongravitational mass-energy of the entire island Universe increase indefinitely. As mass-energy creation at the rate of the Planck power and at the expense of negative gravitational energy is matched by mass-energy dilution via an island Universe's expanding space, so is entropy production matched by entropy dilution. Thus the negative-energy gravitational field of an island Universe is an inexhaustible fuel (positive mass-energy and negative-entropy = negentropy = less-than-maximum-entropy) source. Gravity is a bank that provides an infinite line of credit and never requires repayment [79]. Planck-power input draws on this infinite line of credit [79], which never runs out — indeed which can*not* run out. [Of course, if (positive) nongravitational mass-energy density remains constant, then so must (negative) gravitational energy density, if the balance of zero total energy is to be maintained. Thus Planck-power input if it exists is really equally of positive nongravitational mass-energy and negative gravitational energy simultaneously.]

Additional questions bearing on the Second Law of Thermodynamics will be discussed in Sects. 5–7. In this chapter, whether concerning Planck-power input or otherwise, we limit ourselves to considerations of thwarting the heat death within the restrictions of

the Second Law. Nonetheless we note that the universal validity of the Second Law of Thermodynamics has been seriously questioned [80–84], albeit with the understanding that even if not universally valid at the very least it has a very wide range of validity [80–84].

There are two difficulties that should at least be briefly mentioned and, even if only briefly and only incompletely, also addressed. (i) In order for negative gravitational energy to balance positive mass-energy of a hydrogen atom (or of any other entity), a hydrogen atom (or other entity) newly created via Planck-power input would have to interact gravitationally *infinitely fast or instantaneously* [85,86] — and hence universally simultaneously [85,86] — with our *entire* L-region of the Universe within our cosmological event horizon [3]. But if a *signal* of mass-energy and/or information is not *transmitted*, no violation of relativity is required [85,86]. Perhaps this may be possible if, as suggested in the third paragraph of this Sect. 4, Planck-power input occurs *initially* as positive-cosmological-constant Λ [65–67], with $f \sim 10^{-5}$ thereof, then hopefully, somehow, via an as-yet-unknown mechanism, being transformed into hydrogen. Perhaps the gravitational interaction of positive-cosmological-constant Λ [65–67], and thence of hydrogen atoms (and/or other entities) newly created therefrom via Planck-power input can be instantaneously "*rubber-stamped*" onto our entire L-region at once, rather than being *transmitted* as a "*signal*" from one place to another within our L-region. (ii) *Even if* an interaction, or any other process such as "rubber-stamping," can be infinitely fast or instantaneous — and hence universally simultaneous — it can be so in only *one* reference frame [86]. A superluminal phenomenon, even be it only the motion of a geometric point that possesses no mass-energy and carries no information (for example the intersection point of scissors blades) [86], can be infinitely fast and hence instantaneous — universally simultaneous — in only one reference frame [86] (as a subluminal phenomenon can be infinitely slow — at rest — in only one reference frame [86]).[10] But there is a natural choice for this frame: The comoving frame [7], in which the cosmic background radiation and Hubble flow are isotropic [7], even if not an absolute rest frame, is at least a preferred rest frame [87], indeed *the* preferred rest frame [87], of our L-region of the Universe. If any one reference frame can claim to be preferred, it is the comoving frame [7,87]. Since by the cosmological principle [51] there is nothing special about our L-region of the Universe, the same likewise obtains in any other L-region thereof. The existence of this universal preferred frame [7,87] implies the existence of a preferred, perhaps even absolute, cosmic time τ [4,5,87]. A clock in the comoving frame measures cosmic time τ [4,5,87] — the *longest* possible elapsed time from the Big Bang until now (and also the *longest* possible elapsed time from the Big Bang to the Big Crunch in an oscillating cosmology [16,88–94]) — clocks in all other frames measure *shorter* elapsed times [4,5,88–94].[11] A clock in the comoving frame also measures the *longest* possible elapsed time $\Delta\tau$ corresponding to a given decrease in the temperature of the cosmic

[10] (Re: Entry [86], Ref. [2]) Special Relativity permits arbitrarily fast superluminal phenomena that transmit no mass-energy or information, as well as mutual velocities up to $2c$: see pp. 56 and 70 of Ref. [2]. Section 2.10 of Ref. [2] states that the speed U of transmission of information must not exceed c if violation of causality is to be prevented in Special Relativity. But Eqs. (2.21) and (2.22) in Sect. 2.10 of Ref. [2] at least suggest the possibility that Special Relativity may be consistent with a somewhat less conservative limit, namely $U \leq c^2/v$, where v is the relative velocity between the transmitter and receiver. Of course to guarantee causality Nature must then have a method to checkmate any attempt by the transmitter and/or receiver to "cheat" by increasing v while a signal is en route.

[11] (Re: Entry [88], Ref. [2]) The following is a near-quote from p. 402 of Ref. [2]: "Though unlikely to represent the actual Universe (according to present data) the oscillating-Universe model is interesting in itself."

http://dx.doi.org/10.5772/61642

background radiation (or to a given increase in this temperature during the contracting phase in an oscillating cosmology). This *longest* possible elapsed time *is* cosmic time [4,5,87]. A clock moving at velocity v *relative to the comoving frame* [7,87] measures times *shorter* by a ratio of $\left(1 - v^2/c^2\right)^{1/2}$ [7,87]. Thus the existence of this universal preferred frame and hence of cosmic time [4,5] weakens [87] the concept of relativity of simultaneity [85] as obtains within "the featureless vacuum of Special Relativity" [4,5,85–87]: Events, *even if spatially separated*, can be considered *absolutely* simultaneous if they occur *when* — with "when" having an *absolute* meaning — the cosmic background radiation as observed in the comoving frame has the same temperature, this temperature currently decreasing monotonically with increasing cosmic time τ since the Big Bang [4,5,87].[12] (Simultaneity of *non*-spatially-separated events is absolute even in Special Relativity [85].) Also, the contribution to the total nongravitational mass of our L-region of the Universe of a body of rest-mass [95] m is equal to m only if it is at rest in the comoving frame; if it moves at velocity v relative to the comoving frame [7,87] then its contribution is $m\left(1 - v^2/c^2\right)^{-1/2}$ [7,87]. For a zero-rest-mass particle the contribution is $m = E/c^2$ where E is its energy as measured in the comoving frame (for example $m = E/c^2 = h\nu/c^2$ for a photon of frequency ν as measured in the comoving frame). Thus the total nongravitational mass-energy M_0 of our L-region as per Eq. (1) is that measured *with respect to the comoving frame*.

It should be noted that these two difficulties (i) and (ii) discussed in the immediately preceding paragraph [85,86] also plague Universes created via the Everett interpretation of quantum mechanics [96–98], provided that creation of Everett Universes [96–98] is required to obey mass-energy conservation (no creation of mass-energy *ex nihilo* [21–25]). The creation of Everett Universes [96–98] with no higher entropy (or entropy density if they are infinite) than that of their precursor Universe could perhaps obtain for reasons similar to Planck-power input into our L-region being at positive but less-than-maximum entropy as per our considerations in this Sect. 4 — perhaps most importantly kinetic control winning over thermodynamic control [72–77].

5. The Planck power: One-time and two-time low-entropy boundary conditions, and minimal Boltzmann brains

As discussed in Sects. 3 and 4 (recall especially the third paragraph of Sect. 4), the simplest model of Planck-power input entails a fixed positive cosmological constant Λ. Also, from the viewpoint of General Relativity [65–67], a fixed cosmological constant Λ is the simplest choice for Λ [65–67]. Yet we should also consider other possibilities [66].

False-vacuum high-energy-density-scalar-field regions — the inflaton field — of the Multiverse separating island Universes [1] inflate much faster than they decay to non-inflating true-vacuum regions. Hence while inflation had a beginning once begun it is eternal [99]. Within island Universes high-cosmological-"constant" regions play essentially the same role that inflationary regions play between island Universes: they double in size much faster than their half-life against decay, so each island Universe expands forever, albeit more slowly than inflationary regions separating island Universes [1,100]. Yet the

[12] (Re: Entry [87]) The phrase "the featureless vacuum of Special Relativity" is a quote from a very thoughtful and insightful letter from Dr. Wolfgang Rindler, most probably in the 1990s, in reply to a question that I raised concerning relativity of simultaneity.

cosmological "constant" is not high everywhere in an island Universe [1]; in L-regions and O-regions such as ours regions it is sedate. As decay of the inflaton field gives birth to island Universes, within each island Universe decay of high-cosmological-constant-field regions gives birth to new sedate L-regions and O-regions such as ours. In these sedate L-regions and O-regions, the cosmological "constant" may eventually decay to negative values, resulting in a Big Crunch — and perhaps oscillatory behavior, even as entire island Universes expand forever and the spaces between them expand forever even faster. For simplicity, as noted in the first paragraph of this Sect. 5, we thus far in this chapter (except for brief parenthetical remarks in the second-to-last paragraph of Sect. 4) considered our L-region and O-region to be ever-expanding [more often than not assuming constant $H(\tau) = H_0$ for consistency with a fixed positive cosmological constant Λ and for maximum simplicity]. We now offer a few brief speculations concerning the role of the Planck power if the Universe, or at least our L-region and O-region, is oscillating with two-time low-entropy boundary conditions at the Big Bang and at the Big Crunch [16,88–94,101–105]. It is important to note that there exist oscillating cosmological models, including those with thermodynamic rejuvenation, both in conjunction with and apart from the concept of an inflationary Multiverse [16,88–94,101–105]. Some of these models [16,88,89,101–104] were developed well before inflationary cosmology, when our *observable* Universe or O-region was construed to be the *entire* Universe or at least a major fraction thereof. Within inflationary cosmology, it has been theorized based on quantum considerations that the probability that an oscillating L-region and O-region will have a given lifetime τ_{osc} from Big Bang to Big Crunch decreases towards zero with increasing τ_{osc} such that $\tau_{\text{osc}} = \infty$ — a nonoscillating L-region and O-region — is impossible [88,90], even as entire island Universes expand forever and the spaces between them expand forever even faster. Based on this theoretical analysis [88–90] the dark energy *must* eventually switch sign and become attractive instead of repulsive [88–90]: Hence according to this theoretical analysis [88–90] not only the current acceleration of our L-region's and O-region's expansion but even the expansion itself *must* be a passing fad — our L-region and O-region *must* be oscillatory [88–90].[13] Shortly we will discuss Dr. Roger Penrose's central point [61,62] concerning entropy in the context of both ever-expanding and oscillatory behavior.

If Planck-power input is positive when our L-region expands, could it be negative if and when it contracts? Could this reduce or at least help to reduce the (nongravitational) mass-energy, and hence also entropy, during contraction, possibly to zero, by the time of the Big Crunch? If so, could a singularity at the Big Crunch thereby be evaded, thus ensuring a new thermodynamically fresh Big Bang to begin a new cycle? Moreover, since the Planck power (whether or not divided by c^2) does *not* contain $\hbar$, but only G and c, would or at least might this evading of a Big Crunch singularity be a *classical* process *in*dependent of

[13] (Re: Entries [89]–[94], Refs. [1], [12], and [90]–[93]) The analysis showing that our L-region and O-region must be oscillatory is discussed qualitatively in the passages from Ref. [1] cited in Entry [89], with more technical discussions provided in Ref. [90]. At the 27th Texas Symposium on Relativistic Astrophysics, held at the Fairmont Hotel in Dallas, Texas, December 8–13, 2013, I asked Dr. Michael Turner about the theory discussed in Entry [89] and Ref. [90], according to which the Universe, or at least our L-region and O-region, *must* be oscillatory, also mentioning Entry [89] and Ref. [90]. Dr. Turner is familiar with the passages from Ref. [1] cited in Entry [89] and with Ref. [90], but nevertheless seemed to favor an ever-expanding Universe. An alternative model of an oscillating Universe is discussed in the work in Ref. [63] cited in Entry [92]. In the alternative model of an oscillating Universe investigated in this work, even a Big Rip is shown to be consistent with and indeed part of an oscillating Universe's life cycle. The work in Ref. [63] cited in Entry [93] considers related issues. Also, we should mention that an oscillatory Universe is closer to Einstein's conception of cosmology than a nonoscillatory one. A closed oscillating Universe with $\Lambda = 0$, similar to that considered by Dr. Albert Einstein in the early 1930s, is discussed in the material from Ref. [12] cited in Entry [94].

http://dx.doi.org/10.5772/61642

quantum effects, if not absolutely then at least via opposing quantum effects canceling out, as $\hbar$ cancels out in the division $P_{\text{Planck}} = E_{\text{Planck}}/t_{\text{Planck}}$ [12–15]? Note again that perhaps similar canceling out obtains with respect to the Planck speed $l_{\text{Planck}}/t_{\text{Planck}} = c$: c is the fundamental speed in the classical (nonquantum) theories of Special and General Relativity.

But *negative* Planck-power input requires entropy *reduction*. Hence it seems to require *two-time* low-entropy boundary conditions [101–105] at the Big Bang *and* at the Big Crunch — although two-time, or one-time, low-entropy boundary conditions can also obtain in a "traditional" oscillating Universe without any (positive or negative) Planck-power input [16,88,89,101–105] and without any (repulsive or attractive) dark energy or cosmological constant [16,88,89,101–105]. (In "traditional" oscillating cosmologies, one-time low-entropy boundary conditions imply increasing entropy from cycle to cycle, with each succeeding cycle being longer and reaching a larger maximum size [106,107]. In nonoscillating, ever-expanding, cosmologies, only one-time low-entropy boundary conditions can occur.) Two-time low-entropy boundary conditions require that not only the Big Bang but also the Big Crunch must be special [61,62,101–105]. But even *one-time* low-entropy boundary conditions at the Big Bang that are required for our L-region and O-region to exist as it is currently observed are equally special [61,62]. We will not address the question of whether or not the decrease in entropy during the contracting phase of an oscillating universal cycle imposed by two-time low-entropy boundary conditions [101–105] should be construed as contravening the Second Law of Thermodynamics. It could perhaps be argued that, *within* the restrictions of the Second Law, given two-time low-entropy boundary conditions [101–105] there is no *net* decrease in entropy for an entire cycle, or that two-time low-entropy boundary conditions [101–105] impose such a tight constraint on an oscillating Universe's journey through phase space that there is *no change* in entropy from the initial and final low value during a cycle. In accordance with the third-to-last paragraph of Sect. 4, in ideas developed in this chapter per se (as opposed to brief descriptions of ideas developed in cited references) we limit ourselves to considerations of thwarting the heat death within the restrictions of the Second Law of Thermodynamics. Nonetheless we again note that the universal validity of the second law has been seriously questioned [80–84], albeit with the understanding that even if not universally valid at the very least it has a very wide range of validity [80–84].

The reduction of the (nongravitational) mass-energy of a contracting Universe to zero or at least close to zero at the Big-Crunch/Big-Bang = Big Bounce event might thus be a way, although not necessarily the only way [101–105], to ensure zero entropy — the entropy of nothing is *perforce* zero — or at least low entropy at the Big Bounce. It should be noted that a zero- or at least low-entropy state at the Big Bounce is *imposed* in models with *two-time* low-entropy boundary conditions [101–105]. Thus the cosmic time [4,5] interval from the Big Bang to the Big Crunch can be incomparably shorter than and is totally unrelated to the Poincaré recurrence time [108].[14]

But whether low-entropy or equivalently high-negentropy boundary conditions are one-time, or two-time in oscillating cosmologies [61,62,101–105], Dr. Roger Penrose's central

[14] (Re: Entry [108], Ref. [5]) Contrary to what is stated on p. 192 of Ref. [5], in ever-expanding cosmological models Poincaré fluctuations on the scale of galactic — or smaller, indeed, even minimal-Boltzmann-brain — dimensions in spite of the dissipation due to expansion would *not* be expected, because the energy of starlight and ultimately *all* energy would be irrevocably lost from each and every galaxy into infinitely-expanding space and (without compensating input via a Planck-power or other mechanism, which is not considered on p. 192 of Ref. [5]) never replaced.

point [61,62] concerning entropy survives unscathed. This point had been brought out previously [108–110], but Dr. Roger Penrose's more modern analysis [61,62] takes into consideration inflation, which was not generally recognized prior to the late 1970s [108–110]. (See Sect. 6 concerning the connection with inflation.) This point begins with but does not end with recognizing that the L-region and O-region of our Universe are not merely special. They are *much more* special than they have to be — their negentropy is *much* greater than is required for conscious observers to exist. *By far* the *minimum* negentropy consistent with conscious observation would be that required for the *minimal* existence of a *single minimally*-conscious observer — *one and only one minimal Boltzmann brain* [111–118] with no body or sense organs, and with zero information including zero sensory input even if fictitious [112] and zero memory even if fictitious [113], save only the *minimal* information that one exists and is conscious and even this *minimal* information only for most *minimal* fleeting split-second of conscious existence consistent with recognition that one exists and is conscious, in an otherwise *maximum*-entropy and therefore dead L-region and O-region of our Universe — no other observers, no Sun or other stars, no Earth or other planets, no Darwinian evolution, no nothing (at any rate no nothing worthwhile). Input of *any* sensory information even if fictitious [112], and/or *any* memory even if fictitious [113], is incompatible with the *minimalness* of a Boltzmann brain required by Boltzmann's exponential relation between negentropy $\sigma \equiv S_{\max} - S$ and its associated probability $\text{Prob}\,(\sigma) = \exp\,(-\sigma/k)$. [Note: *Negentropy* $\sigma \equiv S_{\max} - S$ should not be confused with the *entropy change* ΔS associated with a given reaction or process introduced in the paragraph containing Eq. (8).] Even fictitious sensory input [112] or fictitious memory [113], as in a dream or in a simulated Universe, requires larger σ than none at all and hence is exponentially forbidden. Thus Boltzmann's exponential relation $\text{Prob}\,(\sigma) = \exp(-\sigma/k)$ allows not *any* Boltzmann brain but only a *minimal* Boltzmann brain — and only *one* of them. Based *solely* on Boltzmann's exponential relation $\text{Prob}\,(\sigma) = \exp(-\sigma/k)$ a lone *minimal* Boltzmann brain is not merely by far but *exponentially by far* the most probable type of observer to be and *exponentially by far* the most probable type of L-region and O-region of our Universe — or of any Universe in the Multiverse — to find oneself in: One should then expect not even fictitious sensory input [112], not even fictitious memory [113], but only the most fleeting split-second of conscious existence consistent with recognition that one is conscious.

But a basis *solely* on Boltzmann's relation $\text{Prob}\,(\sigma) = \exp(-\sigma/k)$ is incorrect, or at the very least incomplete. Boltzmann's relation $\text{Prob}\,(\sigma) = \exp(-\sigma/k)$ is valid *only* assuming thermodynamic equilibrium — that the ensemble of L-regions and O-regions corresponds to that at thermodynamic equilibrium. Probably the most powerful argument against this being the case is the *vast* disparity between our L-region and O-region that we *actually* observe and what one *would* observe as per the immediately preceding paragraph based *solely* on Boltzmann's relation $\text{Prob}\,(\sigma) = \exp(-\sigma/k)$. This disparity, the minimal-Boltzmann-brain disparity, by a factor of $O \sim 10^{10^{123}}$ [61,62], utterly dwarfs the disparity by a factor in the range of $O \sim 10^{120}$ [119] to $O \sim 10^{123}$ [120] between the observed and predicted values of the cosmological constant [119,120] — indeed it may utterly dwarf *all other disparities combined* [121]. These other disparities [121] relate mainly to the fundamental and effective laws of physics and physical constants requisite for the existence even of a minimal Boltzmann brain. Yet, even apart from viewpoints [122] that not all of them [121] may be significant, they are utterly dwarfed by the minimal-Boltzmann-brain disparity that obtains *even given* these requisite fundamental and effective laws of physics and physical

http://dx.doi.org/10.5772/61642

constants.[15] In contrast to minimal Boltzmann brains, we are sometimes dubbed "ordinary observers" [115–117] — but based *solely* on Boltzmann's relation $\mathrm{Prob}\,(\sigma) = \exp(-\sigma/k)$ dubbing us even as *extra*ordinary observers would be a vast understatement. Indeed the same reasoning can be extended to *extra*ordinary observers. For, based *solely* on Boltzmann's $\mathrm{Prob}\,(\sigma) = \exp(-\sigma/k)$, *exponentially by far* the most probable *extra*ordinary observer (say, a human with a typical life span) is a *minimal extra*ordinary observer, and only *one* of these per L-region or O-region. While σ required for a lone *minimal extra*ordinary observer greatly exceeds that required for a lone minimal Boltzmann brain, it is *still* utterly dwarfed by the actual σ of our L-region and O-region: The disparity of $\mathrm{Prob}\,(\sigma) = \exp(-\sigma/k)$ between that corresponding to a lone *minimal extra*ordinary observer and that corresponding to our observed L-region and O-region is *still* by a factor of $O \sim 10^{10^{123}}$ [61,62]. We are privileged to be not merely *minimal extra*ordinary observers but *super-extra*ordinary observers — more correctly *hyper-extra*ordinary observers — with an entire Universe to explore and enjoy [61,62].

There are many arguments against Boltzmann-brain hypotheses [111–118]. Indeed, if there exist (a) *imposed* one-time low-entropy boundary conditions, (b) *imposed* two-time low-entropy boundary conditions [16,88–94,101–105] in an oscillating L-region and O-region, or (c) Planck-power (or other [21–25,33–35]) *imposed* low-entropy mass-energy input such as hydrogen in a nonoscillating one [78], then such *imposition* would *preclude* thermodynamic equilibrium. Indeed, given (b) or (c), thermodynamic equilibrium would not only be precluded but be precluded *forever*. Given (b) or (c), there would be no need to assume a decaying or finite-lived Universe [117] to help explain consistency with our observations. But even given (a) the heat death $\sigma = 0$ need *not* be the most probable *current* state of the L-region or O-region of our Universe and hence a *minimal* Boltzmann brain [111–118] need *not* be the most probable *current* observer therein, because at the *current* cosmic time decay to maximum entropy has not yet occurred. Since by the cosmological principle [51] our L-region and O-region are nothing special, this must likewise be true with respect to any L-region or O-region in our island Universe — and likewise with respect to those in any other island Universe in the Multiverse. Moreover, it has even been argued that low-entropy boundary conditions are *not* required to avoid minimal Boltzmann brains being exponentially by far the most probable type of observer, or even the most probable type of observer at all [116]. Also, it has been argued that special, i.e., low-entropy, conditions are not required at Big Bangs or Big Bounces [123,124]. [Clustering of matter at $t = 0$, which might typically be expected to increase entropy in the presence of gravity [125,126], does not do so because in this model [123,124] it is prevented owing to positive kinetic energy equaling negative gravitational energy in magnitude, so that the total energy (which in a Newtonian model excludes mass-energy) equals zero. But on pp. 3–4 of Ref. [124], friction, which generates entropy, is invoked during the time evolution of the system. Frictional damping, by degrading part of the macroscopic kinetic energy of any given pair of objects into microscopic kinetic energy (heat), facilitates their settling into a bound Keplerian-orbit state. But because friction thus generates entropy, this may correspond to a hidden, overlooked, pre-friction low-entropy assumption concerning the initial $t = 0$ state of this model [123,124] in either of its two directions of time [123,124]. But a Kepler pair can be formed without friction, for example via a three-body collision wherein a third body removes enough macroscopic

[15] (Re: Entry [122], Ref. [112]) Reference [112] provides discussions of a spectrum of numerous viewpoints concerning Multiverses and related topics, Dr. Steven Weinberg's viewpoint among this spectrum of viewpoints.

kinetic energy from the other two (without degrading any into heat) that they can settle into a bound Keplerian-orbit state.]

Low-entropy Planck-power (or other [21–25,33–35]) input such as hydrogen in nonoscillating cosmologies, or two-time low-entropy boundary conditions in oscillating ones [61,62,101–105], would enable our Universe — and likewise any Universe in the Multiverse — to *forever* thwart the heat death predicted by the Second Law of Thermodynamics. It should be noted that there also are other ways that the heat death can be thwarted: see, for example, Ref. [127]. Hopefully, one way or another, the heat death *is* thwarted in the *real* Universe, whether within an inflationary Multiverse [89–94,105] or otherwise [88,89,101–105,127].

Perhaps we should also note that the fraction $f \sim 10^{-5}$ of Planck-power input as hydrogen mentioned in Sect. 4 would maintain our L-region and O-region *much farther* from thermodynamic equilibrium than is required for existence of one and only one minimal-Boltzmann-brain. Thus *if* Planck-power input exists *then* $f \sim 10^{-5}$ rather than $f = 0$ can*not* be explained owing to our L-region being lucky: Boltzmann's exponential relation $\text{Prob}\,(\sigma) = \exp(-\sigma/k)$ on the one hand, and σ being a monotonically increasing function of f on the other, rules out any values of σ and f larger than the absolute minima that allow the existence of one and only one minimal-Boltzmann-brain obtaining by dumb luck. Thus *if* Planck-power input exists *then* perhaps there is an underlying principle or law of physics *requiring* $f \sim 10^{-5}$ not only in our L-region but in accordance with the cosmological principle [51] in every L-region of our, and also every other, island Universe [1] in the Multiverse [52–58].

6. Dr. Roger Penrose's concerns: Both sides of the inflation issue

We still must consider Dr. Roger Penrose's difficulty with inflation per se, the evidence for inflation not yet being totally beyond doubt [36–50]. Dr. Penrose has shown that, as per Boltzmann's relation between entropy and probability $\text{Prob}\,(\sigma) = \exp(-\sigma/k)$, the probability Prob 1, per "attempt," of creation of a Universe *as far from thermodynamic equilibrium as ours* without inflation, while extremely small, is nevertheless enormously larger than the probability Prob 2 with inflation. That is $\text{Prob}\,2 \lll \text{Prob}\,1 \lll 1$. At the 27th Texas Symposium on Relativistic Astrophysics [8], I asked Dr. Penrose the following question (I have streamlined the wording for this chapter): No matter how much smaller Prob 2 is than Prob 1 (so long as Prob 2, however miniscule even compared to the already miniscule Prob 1, is finitely greater than zero), inflation has to initiate *only once* — after initiating *once* it will then overwhelm all noninflationary regions. Dr. Penrose provided a concise and insightful reply [128], and also suggested that I re-read the relevant sections of his book, "The Road to Reality [15,61,62]" I did so. Dr. Penrose's key argument seems to be centered on squaring inflation with the Second Law of Thermodynamics. Dr. Penrose's central point, already briefly discussed in Sect. 5, begins with but does not end with recognizing that our L-region and O-region are *much* more thermodynamically atypical — with *much* lower entropy — than is required for us to exist even as *hyper-extra*ordinary observers, as opposed to only one of us as a minimal extraordinary observer, let alone only one of us as a minimal Boltzmann brain. Our L-region and O-region are thermodynamically extremely atypical not merely with respect to all possible L-regions and O-regions. They

http://dx.doi.org/10.5772/61642

are thermodynamically extremely atypical even with respect to the extremely tiny subset of already thermodynamically extremely atypical L-regions and O-regions that allow us to exist as *hyper-extra*ordinary observers, as opposed to only one of us as a minimal extraordinary observer, let alone only one of us as a minimal Boltzmann brain. But now the link to inflation per se: As thermodynamically untypical as our L-region and O-region are today, they become as per Boltzmann's $\text{Prob}\,(\sigma) = \exp(-\sigma/k)$ *exponentially ever more* thermodynamically untypical as one considers them backwards in time [61,62]. Thus the disparity *today* by a factor of $O \sim 10^{10^{123}}$ between the minimal-Boltzmann-brain or even minimal-extraordinary-observer hypothesis and observation becomes *exponentially* ever more severe as one considers our L-region and O-region backwards in time [61,62]. Thus the connection with inflation: Since inflation smooths out temperature differences and other nonuniformities, the very existence of temperature differences and other nonuniformities prior to inflation implies lower entropy than without such nonuniformities and hence renders the thermodynamic problem of origins worse not better [61,62]. In fact *exponentially* worse as per Boltzmann's *exponential* diminution $\text{Prob}\,(\sigma) = \exp(-\sigma/k)$ of probability with increasing negentropy ΔS [61,62]. As thermodynamically atypical and hence *exponentially* improbable as our Big Bang was, it must have been thermodynamically *more* atypical and hence *exponentially more* improbable if it was inflation-mediated than if it was not. This is the basic reason for Dr. Penrose's extremely strong inequality $\text{Prob}\,2 \lll \text{Prob}\,1$. (We should, however, cite the remark that prior to inflation there may have been little mass-energy to thermalize [129].) Nevertheless my question still persists: In infinite time, or even in a sufficiently long finite time, even the most improbable event (so long as its probability, however miniscule, is finitely greater than zero) not merely can occur but *must* occur. It has been noted "that whatever physics permitted one Big Bang to occur might well permit many repetitions [130]." But suppose that Universe creations can occur via both noninflationary and inflationary physics. Even if because $\text{Prob}\,2 \lll \text{Prob}\,1$ there first occurred an enormous but finite number N_1 of noninflationary Big Bangs yielding Universes *as far from thermodynamic equilibrium as ours*, so long as Prob 2, however miniscule even compared to the already miniscule Prob 1, is finitely greater than zero, after a sufficiently enormous but finite number N_1 of such noninflationary Universe creations inflation *must* initiate. And it need initiate *only once* to kick-start the inflationary Multiverse. Thereafter the inflationary Multiverse rapidly attains overwhelming dominance over the noninflationary one — with the number N_2 of inflation-mediated Big Bangs yielding Universes *as far from thermodynamic equilibrium as ours* henceforth overwhelming the number N_1 of noninflationary ones by an ever-increasing margin. To reiterate, no matter how much smaller Prob 2 is than Prob 1 (so long as Prob 2, however miniscule even compared to the already miniscule Prob 1, is finitely greater than zero), in infinite time, or even in a sufficiently long finite time, inflation *must* eventually initiate *once*, kick-starting the inflationary Multiverse, which henceforth becomes ever-increasingly overwhelmingly dominant over the noninflationary one. But even if inflation is eternal, it did have a beginning [99], and hence so did the inflationary Multiverse [99].

While in this Sect. 6 the focus is on thermodynamic issues concerning inflation, we note that Dr. Penrose also considers nonthermodynamic issues, specifically the flatness problem [131].

7. Kinetic control versus both heat death and Boltzmann brains?

A tentative solution to the thermodynamic problem of origins, namely dominance of kinetic over thermodynamic control [72–77] has already been proposed, as a reasonable guess, for the special cases of Planck-power input throughout Sect. 4 and Everett-Universe creation in the last paragraph of Sect. 4. We would now like to consider this issue somewhat more generally.

A generalized form of this *prima facie* perhaps reasonable guess might include: (a) Creation in general, by whatever method, both initial via Big Bang with or without inflation, etc. [26–31], via Everett [96–98], and sustained via Planck-power (or other [33–35]) input of equal nonzero quantities of both positive mass-energy and negative gravitational (or other negative [33–35]) energy starting from (zero positive energy + zero negative energy = zero total energy) entails an initial entropy of zero — the entropy of (zero positive energy + zero negative energy = zero total energy) is *perforce* zero. (b) Creation in general, by whatever method, both initial via Big Bang with or without inflation, etc. [26–31], via Everett [96–98], and sustained via Planck-power (or other [33–35]) input of equal nonzero quantities of both positive mass-energy and negative gravitational (or other negative [33–35]) energy starting from (zero positive energy + zero negative energy = zero total energy) is a nonequilibrium process. These processes do not allow enough time for complete thermalization of the input from the initial value of zero entropy of (zero positive energy + zero negative energy = zero total energy) to the maximum possible positive entropy of (nonzero positive energy + nonzero negative energy = zero total energy). Thus even though, *thermodynamically*, *exponentially* the most probable creation, initial or sustained, by any method, would yield a maximum-entropy Universe with *exponentially* the most probable observer a *minimal* Boltzmann brain, *kinetically* the reaction

$$\begin{aligned} &\text{zero positive energy} + \text{zero negative energy} = \text{zero total energy} \\ &\longrightarrow \text{nonzero positive energy} + \text{nonzero negative energy} = \text{zero total energy} \end{aligned} \tag{8 (restated)}$$

occurs too quickly to allow thermodynamic equilibrium = maximum entropy to be attained. Thus creation, initial or sustained, by whatever method, yields (nonzero positive energy + nonzero negative energy = zero total energy) at positive but far lass than maximum entropy, consistently with the Second Law of Thermodynamics but not with the heat death. Thus the basis of our proposed tentative solution to the thermodynamic problem of both initial and sustained-input origins: the reaction (rx) of Eq. (8) is *kinetically* rather than *thermodynamically* controlled [72–77]. This kinetic control does not defeat thermodynamics (specifically the Second Law of Thermodynamics) but it does defeat the heat death. Thus if the reaction of Eq. (8) is kinetically rather than thermodynamically controlled then the heat death is thwarted, but within the restrictions of the Second Law of Thermodynamics. This kinetic as opposed to thermodynamic control could similarly obtain at the initial creation in accordance with Eq. (8) of an oscillating Universe with two-time low-entropy boundary conditions at

http://dx.doi.org/10.5772/61642

the Big Bang and at the Big Crunch [16,61,62,88–94,101–105], and in the case of creation *ex nihilo* [21–25].

But as we discussed in the third paragraph of Sect. 4, perhaps the *simplest* model of Planck-power input is *initially* in the form of the *simplest* possible type of dark energy, corresponding to *positive constant* Λ — a *positive* cosmological *constant* [65–67]. The *simplest* possible type of dark energy, corresponding to *positive constant* Λ — a *positive* cosmological *constant* [65–67] — is perhaps the type of dark energy that is most easily reconcilable with Planck-power input, in particular with *positive constant* Planck-power input. As we have mentioned, it is also simplest with respect to General Relativity [65–67], and it also implies, or at least is consistent with, *constant* $H(\tau) = H_0$ at all cosmic times τ, and hence a fixed size of our L-region, with its boundary (event horizon [2,3]) $R(\tau)$ always fixed at $R_0 = c/H_0$.

Let $\Delta S_{\rm rx}$ be the increase in entropy associated with the reaction (rx) of Eq. (8), with respect to our L-region. If $0 \ll \Delta S_{\rm rx} \ll S_{\rm max} \sim 10^{123}k$, then, on the one hand, the strong inequality $0 \ll \Delta S_{\rm rx}$ ensures an equilibrium constant $K_{\rm eq} = \exp(\Delta S_{\rm rx}/k)$ sufficiently large that the reverse reaction is forbidden for all practical purposes, thus stabilizing creation [72–77]. Thus the strong inequality $0 \ll \Delta S_{\rm rx}$ justifies the placement of only a forward arrow (no reverse arrow) at the beginning of the second line of Eq. (8) [72–77]. On the other hand, the strong inequality $\Delta S_{\rm rx} \ll S_{\rm max} \sim 10^{123}k$ ensures against the doom and gloom that one would dread based *solely* on Boltzmann's relation $\text{Prob}(\Delta S) = \exp(-\Delta S/k)$. Note for example that even if $\Delta S_{\rm rx} = 10^{120}k$ and hence for the reaction (rx) of Eq. (8) $K_{\rm eq} = e^{10^{120}}$, the entropy of our L-region is still only $O \sim 10^{-3}$ of that corresponding to thermodynamic equilibrium and hence still $\sigma \sim 10^{123}k$. [References [73–77] express the equilibrium constant as $K_{\rm eq} = \exp(-\Delta G_{\rm rx}/kT)$, where $\Delta G_{\rm rx}$ is the Gibbs free energy change associated with a reaction in the special case of a system maintained at constant temperature T and constant ambient pressure. (To be precise, the ambient pressure must be maintained strictly constant during a reaction, but the temperature of the reactive system can vary in intermediate states so long as at the very least the initial and final states are at the same temperature, for this definition of $\Delta G_{\rm rx}$ to be valid [132–135].[16] In this special case, $|\Delta G_{\rm rx}|$ is the maximum work that a reaction can yield if $\Delta G_{\rm rx} < 0$ and the minimum work required to enable it if $\Delta G_{\rm rx} > 0$. But in this special case $\Delta G_{\rm rx} = -T\Delta S_{\rm rx}$ where $\Delta S_{\rm rx}$ is the *total* entropy change of the (system + surroundings). Hence $K_{\rm eq} = \exp(-\Delta G_{\rm rx}/kT)$ is the corresponding special case of $K_{\rm eq} = \exp(\Delta S_{\rm rx}/k)$. In this chapter ΔS and $\Delta S_{\rm rx}$ are always taken to be *total* entropy changes of the entire Universe or at least of our L-region thereof.]

[16] (Re: Entry [132], Ref. [132]) One point: On p. 479 of Ref. [132], it is stated that in an adiabatic process all of the energy lost by a system can be converted to work, but that in a nonadiabatic process less than all of the energy lost by a system can be converted to work. But if the entropy of a system undergoing a nonadiabatic process *increases*, then *more* than all of the energy lost by this system can be converted to work, because energy extracted from the surroundings can then also contribute to the work output. In some such cases positive work output can be obtained at the expense of the surroundings even if the change in a system's energy is *zero*, indeed even if a system *gains* energy. Examples: (a) Isothermal expansion of an ideal gas is a thermodynamically spontaneous process, yielding work even though the energy change of the ideal gas is *zero*. (b) Evaporation of water into an unsaturated atmosphere (relative humidity less than 100%) is a thermodynamically spontaneous process, yielding work even though it costs heat, i.e., yielding work even though liquid water *gains* energy in becoming water vapor: see Refs. [133–135] concerning this point.

Thus the doom and gloom that one would dread based *solely* on Boltzmann's relation $\text{Prob}\,(\sigma) = \exp(-\sigma/k)$ does *not* obtain, and furthermore will *never* obtain if there exists *imposed* two-time low-entropy boundary conditions in an oscillating cosmology [16,61,62,88–94,101–105], or Planck-power (or other [21–25,33–35]) *imposed* sustained low-entropy mass-energy input such as hydrogen in a nonoscillating one [78]. Thus creation — initial via Big Bang with or without inflation, etc. [26–31], via Everett [96–98], and sustained via Planck-power (or other [21–25,33–35]) input — being kinetically rather than thermodynamically controlled [72–77] seems to be at least a reasonable tentative explanation of why we are privileged to be not merely *minimal extra*ordinary observers but *super-extra*ordinary observers — more correctly *hyper-extra*ordinary observers — with an entire Universe to explore and enjoy [61,62]. By the cosmological principle [51] we may hope that this is true everywhere in the Multiverse.

As a brief aside, we note that many chemical reactions are similarly kinetically rather than thermodynamically controlled [72–77], in like manner as Eq. (8). While only chemical reactions are discussed in Refs. [72–77], the same principle likewise applies with respect to *all* kinetically rather than thermodynamically controlled processes, for example kinetically rather than thermodynamically controlled physical and nuclear reactions. As we discussed in Sect. 4 if nuclear reactions were thermodynamically rather than kinetically controlled then there would be nothing but (iron + equilibrium blackbody radiation) — an iron-dead Universe.

8. A brief review concerning the Multiverse, and some alternative viewpoints

Four Levels of the Multiverse have been recognized [136–141]: Level I, the infinite number of L-regions and O-regions within an island Universe, with identical fundamental and effective laws of physics but with generally different histories (given the infinite number of L-regions and O-regions per island Universe, identical histories must occur in sufficiently widely separated ones); Level II, an infinite number of island Universes with identical fundamental but different effective laws of physics; Level III, Dr. Hugh Everett's many worlds [96–98]; and Level IV, wherein — within limits [136–142] — different fundamental laws of physics are allowed [136–142].[17]

Dr. Max Tegmark [138,139] writes that Level III is at least in some sense may be equivalent to Levels I+II: Level I incorporates different quantum branches within one single given Hubble volume of an infinity of such volumes contained in an island Universe. Level II incorporates different quantum branches within an entire island Universe. Level III incorporates different Level I and Level II Universes within one single given quantum branch. But it seems that Levels I+II, or at the very least Level I, must exist *first*, because Levels I+II, or at the very least Level I, seems *prerequisite* for the existence of entities capable of executing Dr. Hugh Everett's program [96–98].

[17] (Re: Entry [137], Ref. [112]) Reference [112] provides discussions of a spectrum of numerous viewpoints concerning Multiverses and related topics, Dr. Max Tegmark's viewpoint among this spectrum of viewpoints.

http://dx.doi.org/10.5772/61642

We should note that *if* conscious observers, also referred to as self-aware substructures (SASs) [143–145], are not merely self-aware but also have free will, *then* they have at least *some* degree of *choice* concerning creation of Level III Universes: They then have at least *some* freedom to *choose* whether or not to make a given observation or measurement, which observations and measurements to make, and when to make them. Even if the Everett interpretation [96–98] of quantum mechanics is incorrect [146] and Level III Universes exist only in potentiality until one and only one of them is actualized [146], say via wave-function collapse [147], then an SAS with free will still has this degree of *choice*. Even if the probabilities of the possible outcomes of any given observation or measurement cannot be altered, the set of possible outcomes on offer to Nature depends on which observations and measurements are chosen by an SAS with free will, and when they are on offer depends on when an SAS with free will chooses to observe or measure. Thus irrespective of the character of Level III Universes, *if* free will exists *then* there is this *qualitative* difference between unchosen observations and measurements made by Nature herself, say via decoherence [148,149], and chosen ones made by an SAS with free will. Moreover, "decoherence" is perhaps too strong a term; "delocalization of coherence" seems more correct. Since quantum-mechanical information in general cannot be destroyed, quantum-mechanical coherence in particular is never really destroyed, merely delocalized. As with any delocalization process there is an accompanying increase in entropy. But within a system of finite volume this increase in entropy is limited to a finite maximum value, implying recoherence, or more correctly relocalization of coherence, after a Poincaré recurrence time [108,150,151]. Of course, typical Poincaré recurrence times [108,150,151] of all but very small systems are inconceivably long, but in a very small system at least partial recoherence, or more correctly relocalization of coherence, may occur in a reasonable time. We should note that even before the term "decoherence" had been coined, some aspects of decoherence, or more correctly delocalization of coherence, had been partially anticipated [152,153]. For general reviews concerning the quantum-mechanical measurement problem see, for example, Refs. [149] and [152–155].[18]

Perhaps the concepts considered in this chapter may be at least to some degree applicable to the maximal proposed version of the Multiverse, the Level IV Multiverse [136–145], wherein all well-defined mathematical structures [140–145] — but *not* all arbitrary figments or fantasies of one's imagination [140–142] — would be realized as physically-existing Universes [140–145]. But as Dr. Alex Vilenkin points out, not all mathematical structures, indeed not even all *allowable* mathematical structures given the restrictions stated by Dr. Max Tegmark [140–142], are equal: some are more beautiful and hence more equal than others [156]. Alex Vilenkin writes: "Beautiful mathematics combines simplicity with depth [156]." (But also that "simplicity" and "depth" are almost as difficult to define as "beauty [156]."") But Dr. Alex Vilenkin also writes: "Mathematical beauty may be useful as a guide, but it is hard to imagine that it would suffice to select a unique theory out of

[18] In Chap. 23 of Ref. [152], Dr. David Bohm expresses the viewpoint that classical mechanics should be considered in its own right and as prerequisite for quantum mechanics, rather than as a limiting case of quantum mechanics. This is opposed to the more generally accepted viewpoint that classical mechanics should be considered as a limiting case of quantum mechanics. Moreover, even Dr. David Bohm expresses the latter viewpoint in his own recognition of the Universe as ultimately quantum-mechanical, in Chap. 8 (especially Sects. 8.22–8.23) and Chap. 22 (especially Sects. 22.2–22.3) of Ref. [152]. But, in any case, this is apart from Dr. David Bohm's partial anticipation of certain aspects of decoherence, or more correctly delocalization of coherence, in Sect. 6.12 and Chap. 22 (especially Sects. 22.11–22.12) of Ref. [152].

the infinite number of possibilities [157]." These points are also considered by Dr. Roger Penrose [158]. Yet even so mathematical beauty should have at least *some* selective power. A case in point: Newton's laws have both simplicity and depth, and hence are beautiful. But Einstein's laws have both *greater* simplicity and *greater* depth, and hence are *more* beautiful. The laws of motion have the same form in all reference frames in General Relativity but not in Newton's theory (for example, Newton's theory requires extra terms for centrifugal and Coriolis forces in rotating reference frames), thus General Relativity has greater simplicity; additionally, Newton's theory is a *limiting case* of Einstein's but not vice versa, thus General Relativity also has greater depth. Hence might a Universe wherein Newton's laws are the *fundamental* laws, not merely a limiting case of relativity and quantum mechanics, be denied physical existence in a Level IV Multiverse — because even though it is a beautiful mathematical structure, it is not the *maximally*-beautiful one that *maximally* entails both simplicity and depth? While (even neglecting quantum mechanics) we can*not* be sure if General Relativity *is* the maximally-beautiful mathematical structure, we *can* be sure that Newtonian theory, while beautiful, is *not* maximally beautiful. Moreover, while the Multiverse is eternal, it nonetheless, at least below Level IV [136], did have a beginning [99]. The laws of quantum mechanics — *our* laws of quantum mechanics — governed the initial tunneling event that created not merely our Universe but the Multiverse, at least through Level II [99,136]. Thus these laws, on whatever tablets they are written, must have existed before, and must exist independently of, the Multiverse at least through Level II [99,136] — not merely of our island Universe [99]. Concerning Level III, it seems that Levels I+II, or at the very least Level I, must exist *first*, because Levels I+II, or at the very least Level I, seems *prerequisite* for the existence of entities capable of executing Dr. Hugh Everett's program [96–98]. But might the prerequisites for a beginning and for the pre-existence of *our* laws of quantum mechanics be general, operative even at Level IV [136]? But if so then might Level IV — but not Levels I, II, and III — be more restricted than has been suggested [136]? For then might *our* laws of quantum mechanics be part of the *one* maximally-beautiful mathematical structure that *maximally* entails both simplicity and depth — *our* fundamental (not merely effective) laws of physics [136] — after all? Then perhaps this *one* maximally beautiful mathematical structure, this *maximal* possible entailment of both simplicity and depth, is the only one realized via physically-existing Universes. But if this is the case then the question arises: Why does this *one* maximally beautiful mathematical structure permit life [159] (at the very least, carbon-based life as we know it on Earth)?

We must admit that in this chapter we have not even scratched the surface, as per this paragraph and the two immediately following ones. There are many alternative viewpoints concerning the Multiverse and related issues. We should at least mention a few of them that we have not mentioned until now. According to at least one of these viewpoints, inflation is eternal into the past as well as into the future, and hence has no beginning as well as no end [160–162]. But perhaps this is compatible with inflation having a beginning if regions of inflation in the forward and backward time directions are disjoint and incapable of any interaction with each other [163]. Then perhaps observers in both types of regions would consider their home region to be evolving forward, not backward, in time. According to other viewpoints, inflation not only has a beginning but also has an end — eternal inflation is impossible [164,165]. According to one of these viewpoints, the end of inflation is imposed by the increasingly fractal nature of spacetime [164,165]. We also note that Dr. Roger Penrose considered another difficulty associated with possible fractal nature of spacetime: inflation does not solve the smoothness and flatness problems if the structure of spacetime is fractal,

http://dx.doi.org/10.5772/61642

or worse than fractal [166]. According to another of these viewpoints, the end of inflation is imposed by the Big Snap, according to which expansion of space will eventually dilute the number of degrees of freedom per any unit volume, and specifically per Hubble volume, to less than one, although the Universe will probably be in trouble well before the number of degrees of freedom per Hubble volume is reduced to one [167,168]. But perhaps new degrees of freedom can be created to compensate [167,168]. Perhaps Planck-power input, if it exists, can, because it replenishes mass-energy, also replenish degrees of freedom — thereby precluding the Big Snap. In Dr. Max Tegmark's rubber-band analogy, this corresponds to new molecules of rubber being created as the rubber band stretches, thereby keeping the density of rubber constant [167,168]. But, with or without a Big Snap [167,168], if inflation does have an end for any reason whatsoever, then my question to Dr. Roger Penrose in Sect. 6 is answered negatively.

There are also many proposed solutions to the entropy problem (why there is so very much more than one minimal Boltzmann brain in our L-region and O-region), some of which we have already discussed and/or cited in Sects. 5–8, other than Planck-power input. But there are still other proposed solutions to the entropy problem. One other proposed solution that we have not yet cited entails quantum fluctuations ensuring that every baby Universe starts out with an unstable large cosmological-constant, which corresponds to low total entropy because it is thermodynamically favorable for the consequent high-energy false vacuum to decay spontaneously [169,170]. Yet another proposed solution that we have not yet cited entails observer-assisted low entropy [168].

There are also many alternative viewpoints concerning fine-tuning and life in the Universe. It has been noted that since physical parameters such as constants of nature, strengths of forces, masses of elementary particles, etc., all have real-number values, and the range of real numbers is infinite, then if the probability of occurrence of a given real-number value of a given parameter is uniform, or at least non-convergent, then there is only an infinitessimal probability of this value being within any finite range [171]. But if thee are an infinite number of L-regions and O-regions, this infinity may be as large or even larger. We should also note that while some scientists are favorable towards the idea of fine tuning [172], others are sceptical to the point of not requiring a Multiverse to explain it away, but stating that it is an invalid concept even if our O-region constituted the entire Universe [173–175]. Even this sceptical viewpoint admits that only a very small range of parameter space is consistent with carbon-based life as we know it on Earth [], but assumes that a much larger range of parameter space is consistent with life in general [174]. But life, at least chemically-based life, probably must be based on carbon, because no other element even comes close to matching carbon's ability to form highly complex, information-rich molecules. Even carbon's closest competitor, silicon, falls woefully short. Also, nucleosynthesis in stars forms carbon more easily that silicon [176], so carbon is more abundant [176].

Acknowledgments

I am especially grateful to Dr. Wolfgang Rindler, Dr. Donald H. Kobe, Dr. Bruce N. Miller, and Dr. Roger Penrose for very helpful, thoughtful, and insightful discussions, communications, and advice concerning relativity, cosmology, and thermodynamics. I am also grateful to Dr. Roger Penrose for valuable insights and clarifications concerning

the relation between thermodynamics on the one hand, and inflation and cosmology on the other, and to Dr. Michael Turner for valuable insights and clarifications concerning oscillating versus nonoscillating universes, both at the 27th Texas Symposium on Relativistic Astrophysics, held at the Fairmont Hotel in Dallas, Texas on December 8–13, 2013 (website: nsm.utdallas.edu/texas2013/). I am also grateful to Dr. Wolfgang Rindler, Dr. Bruce N. Miller, and Dr. Roger Penrose for helpful general discussions concerning physics and astrophysics, both at the 27th Texas Symposium on Relativistic Astrophysics and otherwise, and to Dr. Donald H. Kobe for such discussions on various occasions. I also thank Dr. Marlan O. Scully and Dr. Donald H. Kobe for helpful insights concerning decoherence. I am thankful to Dr. Paolo Grigolini for very helpful and thoughtful considerations concerning both earlier and the most recent versions of this manuscript in Special Problems courses. Also, I thank Dr. S. Mort Zimmerman for engaging in general scientific discussions over many years, and both Dan Zimmerman and Dr. Kurt W. Hess for brief yet helpful discussions concerning this chapter and for engaging in general scientific discussions at times. I also thank Dr. Iva Simcic, Publishing Process Manager, for much very helpful advice in preparing this chapter and for much extra time to prepare it, and Technical Support at MacKichan Software for their very helpful advice concerning Scientific WorkPlace 5.5.

Author details

Jack Denur

Address all correspondence to: jackdenur@my.unt.edu

Electric & Gas Technology, Inc., Rowlett, Texas, USA

References

[1] Vilenkin A. Many Worlds in One: The Search for Other Universes. New York: Hall and Wang; 2007. DOI: 10.1063/1.2743129. See Chaps. 10 and 11, pp. 203–205, and the associated Notes including references cited therein.

[2] Rindler W. Relativity: Special, General, and Cosmological. 2nd ed. New York: Oxford University Press; 2006, pp. 367, 374, and 384.

[3] Reference [2], Sects. 16.1G and 17.1–17.4 (especially Sect. 17.3), also Exercise 16.3 on p. 369.

[4] Reference [2], Sects. 16.2–16.4 (especially pp. 359–360 and 367).

[5] Davies PCW. The Physics of Time Asymmetry. 2nd ed. Berkeley: University of California Press; 1977, Sect. 1.4.

[6] Reference [2], pp. 376–377.

[7] Reference [2], Exercise 7.22 on pp. 160–161, Sects. 16.1E–16.1G, p. 362, p. 366, Sect. 17.2, and Exercise 17.4 on p. 388.

[8] 27th Texas Symposium on Relativistic Astrophysics, held at the Fairmont Hotel in Dallas, Texas, December 8–13, 2013. [Internet]. 2013. Available from: nsm.utdallas.edu/texas2013/ [Accessed 2015-12-05]

[10] Lang KR. Essential Astrophysics. Berlin: Springer-Verlag; 2013, Sect. 14.3 and references cited therein.

[11] Planck Collaboration. *Planck* 2015 results. XIII. Cosmological parameters. arXiv:1502.01589v2 [astro-ph.CO] 6 Feb 2015: 67 pages. [Internet]. 2015. Available from: xxx.lanl.gov/abs/1502.01589v2 or arxiv.org/abs/1502.01589v2 [Accessed: 2015-12-05]. See Sects. 5.4 and 5.6.

[12] Misner CW, Thorne KS, Wheeler JA. Gravitation. New York: Oxford; 1973, pp. 10–13 and 1180, and Sects. 43.4, 44.3, and 44.6 (especially pp. 1215–1217).

[13] Planck M. Über irreversible Strahlangdvorgänge. Sitzungsber. Deut. Akad. Wiss Berlin Kl. Math–Phys. Tech. 1899; 440–480.

[14] Bradley W. Carroll BW, Ostlie DA. An Introduction to Modern Astrophysics. 2nd ed. San Francisco: Pearson Addison Wesley; 2007, pp. 1233–1235.

[15] Penrose R. The Road to Reality: A Complete Guide to the Laws of the Universe. New York: Alfred A. Knopf; 2005, Sects. 27.3, 27.10, and 31.1. See also references cited in Sect. 31.1.

[16] Reference [2], Sects. 16.1G–16.1K.

[17] Reference [2], Sects. 16.4, 18.3, and 18.4.

[18] Reference [10], Sects. 14.6 and 15.6.2.

[19] Reference [11], Sect. 6.2.4.

[20] Reference [14], Chap. 29.

[21] Reference [1], pp. 27–28, 37, and 170–171.

[22] Reference [2], pp. 359, 368–369, 387–388, and 411.

[23] Reference [14], pp. 1163–1167.

[24] Harrison ER. Mining Energy in an Expanding Universe. Astrophys. Jour. 1995; 446: 63–66. DOI: 10.1086/175767

[25] Sheehan DP, Kriss VG: Energy Emission by Quantum Systems in an Expanding FRW Metric. arXiv:astro-ph/0411299v1. 11 Nov 2004. 12 pages. [Internet]. 2004. Available from: xxx.lanl.gov/abs/astro-ph/0411299 or arxiv.org/abs/astro-ph/0411299 [Accessed: 2015-12-05]

[26] Tryon EP. Is the Universe a Vacuum Fluctuation? Nature 1973; **246**: 396–397. DOI: 10.1038/246396a0

[27] Reference [1], pp. 11–12 and 183–186.

[28] Filippenko A, Pasachoff JM. A Universe From Nothing. 2 pages. [Internet]. 2001. Available from: https:www.astrosociety.org/publications/a-universe-from-nothing [Accessed: 2015-12-05]

[29] Potter F, Jargodski C. Mad About Modern Physics. Hoboken, NJ; 2005, Question 224, "The Total Energy," on p. 115, and Answer 224, "The Total Energy," along with cited references, on pp. 275–276.

[30] Berman MS. On the Zero-Energy Universe. Int. J. Theor. Phys. 2009; **48**: 3278–3286. DOI: 10.1007/s10773-009-0125-8

[31] Berman MS, Trevisan LA. On Creation of Universe Out of Nothing. Int. J. Modern Phys. B. 2010; **19**: 1309–1313. DOI: 10.1142/S0217827181007342

[32] Reference [15], Sects. 19.7–19.8 (especially Sect. 19.8 and most especially pp. 468–469). See also references cited therein; also Notes for Sects. 19.7–19.8 on pp. 469–470. Of these references, see especially the references cited in Note 19.17 on p. 470.

[33] Reference [5], Sect. 7.2 and references cited therein.

[34] Reference [5], the last sentence on p. 187 and the reference cited therein.

[35] Davies PCW. Is the Universe Transparent or Opaque? J. Phys. A: Math. Gen. 1972; **5**: 1722–1737. DOI: 10.1088/0305-4470/5/12/012. See especially Sect. 8.

[36] Reference [1], pp. 48–52 and 67–69; also Notes for Chaps. 5 and 6 on pp. 211–212, especially Notes 2 and 3 for Chap. 5 on p. 211.

[37] Reference [2], pp. 379–380 and Sects. 18.5–18.6.

[38] Reference [14], Sect. 30.1.

[39] Reference [15], Chap. 28, including Notes for Chap. 28 on pp. 778–781.

[40] Reference [10], pp. 523–524.

[41] Crockett C. Primordial gravitational waves found: Researchers see traces of cosmic expansion just after Big Bang. Science News. April 5, 2014; **185 (7)**.

[42] Krauss LM. Beacon from the Big Bang. Sci. Amer. October, 2014; **311 (4)**: 58–67.

[43] Flauger R, Hill JC, Spergel DN. Toward an understanding of foreground emission in the BICEP2 region. Journal of Cosmology and Astroparticle Physics. 2014; **08**: 039. DOI: 10.1088/1475-7516/2014/08/039

[44] Crockett C. Gravitational wave discovery gives way to dust. Science News. October 18, 2014; **186 (8)**: 7.

[45] Crockett C. Dust obscures possible gravitational wave discovery. Science News. December 27, 2014; **186 (13)**: 16.

[46] Grant A. Gravitational wave claim bites dust. Science News. February 21, 2015; **187 (4)**: 13.

[47] Grant A. The past according to Planck: Cosmologists got a lot right. Science News. March 21, 2015; **187 (6)**: 7.

[48] Milne PA, Foley RJ, Brown PJ, Narayan G. The Changing Fractions of Type 1A Supernova NUV-Optical Subclasses with Redshift. Astrophys. Jour. 2015; **803:20**: 15 pages. DOI: 10.1088/0004-637X/803/120

[49] Reference [15], Sects. 28.4–28.7. See also references cited therein; also Notes for Sects. 28.4–28.7 on pp. 779–781; also pp. 1037–1038.

[50] Witze A. Inflation on trial. Science News. July 28, 2012; **182 (2)**: 20–21.

[51] Reference [2], Sects. 1.11 and 16.1–16.2.

[52] Reference [14], Sect. 13.3 and pp. 886–887, 957–958, 986–987, 1011–1013, 1018–1020, and 1028.

[53] Reference [14], pp. 1011–1013, 1018–1020, and 1028.

[54] Reference [10], Sect. 15.4.2 on pp. 542–545 and references cited therein.

[55] Reference [1], Chap. 10. Also see Notes for Chap. 10. on p. 213 and the reference cited in Note 1.

[56] Reference [14], Chaps. 7–16.

[57] Reference [10], Chaps. 2 and 8–13.

[58] Reference [1], Chap. 10, especially pp. 93–94. Also see Notes for Chap. 10. on p. 213 and the reference cited in Note 1.

[59] Reference [1], pp. 25–28 and 171–175, also Note 4 for Chap. 3 on p. 210.

[60] Reference [5]. See especially Chaps. 1, 2, 4, and 7.

[61] Reference [15], Sects. 19.7–19.8, Chaps. 27 and 28, and references cited therein; also Notes for Sects. 19.7–19.8 and Chaps. 27 and 28 on pp. 469–470, 732–734, and 778–781, respectively; also pp. 1037–1038. See especially Sects. 27.13 and 28.1–28.5, and the associated Notes.

[62] Reference [2], p. 380.

[63] Mersini-Houghton L, Vaas R, Editors. The Arrows of Time: A Debate in Cosmology. Berlin: Springer-Verlag; 2012.

[64] See, for example, Ref. [2], Sect. 16.1D.

[65] Reference [2], Sects. 10.5, 14.2–14.3 and 18.2E. See also the reference cited in Sects. 14.2–14.3.

[66] Reference [2], Sect. 18.2E.

[67] Lovelock D. The Four-dimensionality of Space and the Einstein Tensor. Jour. Math. Phys. 1972; **13**: 874–876. DOI: 10.1063/1.1666069

[68] Reference [11], Sect. 6.3.

[69] Planck Collaboration. *Planck* 2015 results. XIV. Dark energy and modified gravity. arXiv:1502.01590v1 [astro-ph.CO] 5 Feb 2015: 64 pages. [Internet]. 2015. Available from: xxx.lanl.gov/abs/1502.01590 or arxiv.org/abs/1502.01590 [Accessed: 2015-12-05]

[70] Reference [10], Sect. 15.2. See especially Subsects. 15.2.3–15.2.4, and most especially Table 15.1.

[71] Planck Collaboration. *Planck* 2015 results. I. Overview of products and scientific results. arXiv:1502.01582v2 [astro-ph.CO] 10 Aug 2015: 40 pages. [Internet]. 2015. Available from: xxx.lanl.gov/abs/1502.01582v2 or arxiv.org/abs/1502.01582v2 [Accessed: 2015-12-05]. See Fig. 9 and its caption on p. 19.

[72] Noller CR. Chemistry of Organic Compounds, 2nd ed. Philadelphia: W. B. Saunders; 1957, pp. 41–42 and 165–170.

[73] Noller CR. Chemistry of Organic Compounds, 3rd ed. Philadelphia: W. B. Saunders; 1965, Chap. 3.

[74] Noller CR. Textbook of Organic Chemistry, 3rd ed. Philadelphia: W. B. Saunders; 1966, Chap. 3.

[75] Morrison RT, Boyd RN. Organic Chemistry, 6th ed. Englewood Cliffs, N. J.: W. B. Prentice Hall; 1992.

[76] Mahan BM. University Chemistry, Reading, Mass.: Addison Wesley; 1965, Chaps. 8–9.

[77] Mahan BM, Myers RJ. University Chemistry, 4th ed. Menlo Park, Calif.: Bejamin/Cummings; 1980, Chaps. 8–9.

[78] Redd NT. Rivers of Hydrogen Gas May Fuel Spiral Galaxies [Internet]. 2014. Available from: http://www.space.com/24780-spiral-galaxies-hydrogen-gas-river.html (February 21, 2014, 05:25 PM ET) [Accessed 2015-12-05]

[79] Dan Zimmerman, private communications, 2014. When I mentioned this analogy to Dan Zimmerman, he encouraged me to include it in this chapter.

[80] Sheehan, DP, editor. Quantum Limits to the Second Law, AIP Conference Proceedings Volume 643. Melville, N. Y.: American Institute of Physics; 2002.

[81] Nukulov AV, Sheehan DP, editors. Special Issue: Quantum Limits to the Second Law of Thermodynamics. Entropy. March 2004: Vol. 6, Issue 1.

[82] Čápek V, Sheehan DP. Challenges to the Second Law of Thermodynamics: Theory and Experiment. Dordrecht, The Netherlands: Springer; 2005.

[83] Sheehan, DP, editor. Special Issue: The Second Law of Thermodynamics: Foundations and Status. Found. Phys. December 2007: Vol. 37, Issue 12.

[84] Sheehan, DP, editor. Second Law of Thermodynamics: Status and Challenges, AIP Conference Proceedings Volume 1411. Melville, N. Y.: American Institute of Physics; 2011.

[85] See, for example, Ref. [2], Sects. 2.4–2.10 and 3.5, especially Sects. 2.4–2.6 and pp. 56–57. Relativity of simultaneity is most directly discussed in Sects. 2.4–2.6.

[86] Reference [2], Sects. 2.10 and 3.6.

[87] Dr. Wolfgang Rindler, private communications, 1980s–2010s, including at the 27th Texas Symposium on Relativistic Astrophysics, held at the Fairmont Hotel in Dallas, Texas, December 8–13, 2013. [Internet]. 2013. Available from: nsm.utdallas.edu/texas2013/ [Accessed 2015-12-05]

[88] Reference [2], pp. 398 and 402–403.

[89] Reference [1]. See pp. 120–121 and Note 4 for Chap. 16 on p. 219 for pre-inflationary ideas concerning oscillating cosmologies, and Chap. 18 (especially pp. 197–198), Note 3 for Chap. 18 on p. 221, pp. 203–205, and Note 9 for Chap. 19 on p. 222 for oscillating cosmologies considered in light of inflation.

[90] Vilenkin A, Garriga J. Testable anthropic predictions for dark energy. Phys. Rev. D 2003; **67**: 043503–1-11.

[91] Dr. Michael Turner. private communications, at the 27th Texas Symposium on Relativistic Astrophysics, held at the Fairmont Hotel in Dallas, Texas, December 8–13, 2013. [Internet]. 2013. Available from: nsm.utdallas.edu/texas2013/ [Accessed 2015-12-05]

[92] An alternative model of an oscillating Universe is discussed in Freese K, Brown MG, Kinney WH. The Phantom Bounce: A New Proposal for an Oscillating Cosmology. In Ref. [63], pp. 149–156.

[93] Zeh HD. Open Questions Regarding the Arrow of Time. In Ref. [63], pp. 205–217.

[94] Reference [12], Sect. 27.10 (especially Box 27.4 on p. 738).

[95] Reference [2], Sects. 6.2–6.3.

[96] Reference [1], pp. 114–116 and 187–188.

[97] Tegmark M. Our Mathematical Universe. New York: Vintage Books; 2015, p. 179, Chap. 8, pp. 284–286 and 314–315.

[98] Tegmark M. The Mathematical Universe. Found Phys. 2008; 38: 101–150. DOI: 10.1007/s10701-007-9186-9. See Sect. 5.2.

[99] Reference [1], Chaps. 5–8 (especially Chaps. 6 and 8), Chaps. 16 and 17, and pp. 203–205. Also see Notes for Chaps. 5, 6, 8, 16, and 17 and for pp. 203–205 including references cited therein on pp. 211–212 and 219–222. See especially Chaps. 16 and 17, and most especially pp. 180–181 and 204–205.

[100] Reference [1], Chaps. 12 and 13, and Notes for Chaps. 12 and 13 including references cited therein on pp. 214–216.

[101] Reference [5], Chap. 7, especially Sect. 7.4, and most especially the first complete paragraph on p. 193 and the references cited therein, especially Refs. [102–104] of this chapter immediately following.

[102] Cocke J. Statistical Time Symmetry and Two-Time Boundary Conditions in Physics and Cosmology. Phys. Rev. 1967; **160**: 1165–1170. DOI: 10.1103/PhysRev.160.1165

[103] Schmidt H. Model of an oscillating cosmos which rejuvenates during contraction. J. Math. Phys. 1966; **7**: 494–509. DOI: 10.1063/1.1704949

[104] Davies PCW. Closed Time as an Explanation of the Black Body Background Radiation. Nature Physical Science. 1972; **240**: 3–5. DOI: 10.1103/physci240003a0

[105] Hartle J, Hertog T. Arrows of Time in the Bouncing Universes of the no-boundary quantum state. Phys. Rev. D. 2012; **85**: 13 pages. DOI: 10.1038/PhysRevD.85.103524

[106] Reference [5], Sect. 7.3.

[107] Tolman RC. Relativity, Thermodynamics, and Cosmology. Oxford, England: Oxford University Press; 1934. Unaltered and unabridged republication: New York: Dover; 1987, Chaps. IX–X, especially Sects. 130–131, and 169–175 (most especially Sects. 131 and 175).

[108] The Poincaré recurrence time is discussed in Ref. [5], Chap. 3, Sect. 5.2, pp. 131, 144, 164, 173–175, and 192–193, and Sect. 7.4. See also references cited therein.

[109] Reference [5], p. 103.

[110] Reference [2], p. 380 and Sects. 18.5–18.6.

[111] Reference [63]. Boltzmann-brain hypotheses and closely related topics are reviewed on pp. 3, 12–22, 29, 98–102, 193–195, and 205–211. See also relevant references cited therein.

[112] Carr B, Editor. Universe or Multiverse? Cambridge, U. K.: Cambridge University Press; 2007, p. 67.

[113] Reference [97], pp. 305–308 and 313–314.

[114] Reference [98], Sect. 4.2.4.

[115] Bousso R. Vacuum Structure and the Arrow of Time. Phys. Rev. D. 2012; **86**: 16 pages. DOI: 10.1103.PhysRevD.86.123509

[116] Garriga J, Vilenkin A. Watchers of the Multiverse. Journal of Cosmology and Astroparticle Physics. 2013; **05**: 037. DOI: 10.1088/1475-7516/2013/05/037

[117] Page DN. Is the Universe Decaying at an Astronomical Rate? Phys. Lett. B. 2008; **669**: 197–200. DOI: 10.1016/j.physletb.2008.039

[118] Susskind L. Fractal-Flows and Time's Arrow. arXiv:1203.6440v2 [hep-th] 7 Apr 2012. 24 pages. [Internet]. 2012. Available from: xxx.lanl.gov/abs/1203.6440v2 or arxiv.org/abs/1203.6440v2 [Accessed: 2015-12-05]

[119] Reference [1], Chap. 12, and pp. 135–136 and 164.

[120] Reference [97], pp. 140–141, 353–354, and 362–363.

[121] Reference [97], pp. 132–150 and 311–312.

[122] Reference [112]. See especially Chaps. 1, 2, 3, 5, 22, 23, and 25; most especially Chap. 2: Weinberg, S. Living in the Multiverse.

[123] Grant A. Time's Arrow: Maybe Gravity Shapes the Universe Into Two Opposing Futures. Science News. July 25, 2015; **188 (2)**: 15–18.

[124] Barbour J, Koslowski T, Flavio M. Identification of a Gravitational Arrow of Time. Phys. Rev. Lett. 2014; **113**: 118101 (5 pages).

[125] Reference [5], Sect. 4.6.

[126] Vaas R. Time After Time — The Big Bang Cosmology and the Arrows of Time. In Ref. [63], pp. 5–42. See especially pp. 8–9, including Figure 1 on p. 9.

[127] Reference [1], pp. 171–172 and Note 5 for Chap. 16 on p. 219; also, Refs. [24], [25], and [59–63] of this chapter.

[128] Dr. Roger Penrose. private communications, at the 27th Texas Symposium on Relativistic Astrophysics, held at the Fairmont Hotel in Dallas, Texas, December 8–13, 2013. [Internet]. 2013. Available from: nsm.utdallas.edu/texas2013/ [Accessed 2015-12-05]

[129] Reference [2], the last paragraph on p. 413.

[130] Reference [2], p. 416.

[131] Reference [15], p. 756.

[132] Berry RS, Rice SA, Ross J. Physical Chemistry. 2nd ed. New York: Oxford University Press; 2000, pp. 476–479.

[133] Bachhuber C. Energy from the evaporation of water. Am. J. Phys. 1983; **51**: 259–264.

[134] Güémez J, Valiente B, Fiolhais C, Fiolhais M. Experiments with the drinking bird, Am J. Phys. 2003; **71**: 1257–1264.

[135] Temming M. Water, Water Everywhere. Sci. Am. 2015; **313 (3)**: 26.

[136] Reference [97], pp. 134–140, Chap. 12, and pp. 358–370. Concise summaries of Multiverses are provided in Table 6.1 on p. 139, Figure 12.2 on p. 322, and Figure 13.1 on p. 358.

[137] Tegmark M. The Multiverse Hierarchy (Chap. 7). In Ref. [112]. Carr B, Editor. Universe or Multiverse? Cambridge, U. K.: Cambridge University Press; 2007.

[138] Reference [97], pp. 220–226.

[139] Tegmark M. Sect. 7.4. In Ref. [112].

[140] Reference [97]. See especially Chaps. 6, 8, and 10–12, also pp. 358–370 (most especially Chaps. 10 and 12).

[141] Tegmark M. The Multiverse Hierarchy (Chap. 7). In Ref. [112]. See especially Sects. 7.4.4–7.6.

[142] Reference [98], Sect. 5.4.

[143] Reference [97], Chaps. 8 and 11, especially pp. 291–299.

[144] Tegmark M. pp. 116–120. In Ref. [112].

[145] Reference [98], Sects. 2.3, 4.2.4, and 5.3.

[146] Lindley D. Where Does the Weirdness Go?: Why Quantum Mechanics is Strange, But Not as Strange as You Think. New York: Basic Books; 1996, pp. 107–111 and the 2nd Note on p. 233.

[147] Reference [97], pp. 175–183 and Chap. 8.

[148] Reference [97], Chap. 8.

[149] Reference [146], Act III, also Notes for Act III on pp. 234–240.

[150] Reference [5], pp. 173–175.

[151] Reference [146], pp. 177–203, especially pp. 199–203; also the last 3 Notes on p. 237 (the 3rd of these Notes continuing on p. 238), and the 2 Notes on p. 240.

[152] Bohm D. Quantum Theory. Englewood Cliffs, N. J.: Prentice-Hall; 1951. Unaltered and unabridged republication: New York: Dover; 1989, Sect. 6.12 and Chap. 22 (especially Sects. 22.11–22.12).

[153] Reference [5], Sect. 6.3.

[154] Reference [15], Chaps. 29 and 30.

[155] Reference [97], Chaps. 7 and 8, especially Chap. 8.

[156] Reference [1], pp. 202–203 and Notes 6–9 for Chap. 19 on p. 222. See also the reference cited in Note 6.

[157] Reference [1], pp. 201–202 and Notes 2–6 for Chap. 19 on p. 222. See also references cited in Notes 3, 4, and 6.

[158] Reference [15], Sect. 34.9.

[159] Reference [2], Sect. 18.6.

[160] Aguirre A, Gratton S. Steady-state eternal inflation. Phys. Rev. D. 2002; **65**: 083507, 6 pages. DOI: 10.1103/PhysRevD.65.083507

[161] Aguirre A, Gratton S. Inflation without a beginning: A null boundary proposal. Phys. Rev. D. 2003; **67**: 083515, 16 pages. DOI: 10.1103/PhysRevD.67.083515

[162] Aguirre A. Eternal Inflation, past and future. arXiv:0712.0571v1 [hep-th] 4 Dec 2007: 38 pages. [Internet]. 2007. Available from: xxx.lanl.gov/abs/0712.0571 or arxiv.org/abs/0712.0571 [Accessed: 2015-12-14]

[163] Vilenkin A. Arrows of time and the beginning of the universe. Phys. Rev. D. 2013; **88**: 043516, 10 pages. DOI: 10.1103/PhysRevD.88.043516

[164] Mersini-Houghton L, Perry MJ. The end of eternal inflation. Class. Quantum Grav. 2014; **31**: 165005, 11 pages. DOI: 10.1088/0264-9381/31/16/165005

[165] Mersini-Houghton L, Perry MJ. Localization on the landscape and eternal inflation. Class. Quantum Grav. 2014; **31**: 215008, 17 pages. DOI: 10.1088/0264-9381/31/21/215008

[166] Reference [15], Sects. 28.4–28.5. See especially pp. 746–747 and 756–757.

[167] Reference [97], pp. 366–370.

[168] Tegmark M. How unitary cosmology generalizes thermodynamics and solves the inflationary entropy problem. Phys. Rev. D. 2012; **85**: 123517, 19 pages. DOI: 10.1103/PhysRevD.85.123517

[169] Carroll SM, Chen J. Spontaneous Inflation and the Origin of the Arrow of Time. arXiv:0410270v1 [hep-th] 27 Oct 2004: 36 pages. [Internet]. 2004. Available from: xxx.lanl.gov/abs/0410270 or arxiv.org/abs/0410270 [Accessed: 2015-12-15]

[170] Carroll SM. From Eternity to Here: The Quest for the Ultimate Theory of Time (Kindle Edition, Version 6). New York: Penguin; 2010. See especially Chaps. 15–16 and most especially Locations 6576–6846; also, Notes 291–295 at Locations 8218–8233.

[171] Davies P. Universes galore: where will it all end? (Chap. 28). In Ref. [112]. Carr B, Editor. Universe or Multiverse? Cambridge, U. K.: Cambridge University Press; 2007. See especially Sect. 28.3.2.

[172] Barnes LA. The Fine-Tuning of the Universe for Intelligent Life. Publications of the Astronomical Society of Australia. 2012; **29**: 529–564. DOI: 10.1071/AS12015

[173] Stenger VJ. The Fallacy of Fine-Tuning: Why the Universe Is Not Designed for Us (Kindle Edition). Amherst, N. Y.: Prometheus Books; 2011.

[174] Stenger VJ. God and the Multiverse. Amherst, N. Y.: Prometheus Books; 2014. See especially Chap. 16.

[175] Stenger VJ. Defending The Fallacy of Fine-Tuning. arXiv:1202.4359v1 [physics.pop-ph] 28 Jan 2012: 12 pages. [Internet]. 2012. Available from: xxx.lanl.gov/abs/1202.4359 or arxiv.org/abs/1202.4359 [Accessed: 2015-12-14]

[176] See the Wikipedia article entitled "Abundance of the chemical elements." [Internet]. 2015. Available from: www.wikipedia.org [Accessed: 2015-12-16]

Chapter 10

Absolute Zero and Even Colder?

Jack Denur

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/61641

Abstract

We consider first the absolute zero of temperature and then negative Kelvin temperatures. The unattainability formulation of the Third Law of Thermodynamics is briefly reviewed. It puts limitations on the quest for absolute zero, and in its strongest mode forbids the attainment of absolute zero by any method whatsoever. But typically it is stated principally with respect to thermal-entropy-reduction refrigeration (TSRR). TSRR entails reduction of a refrigerated system's thermal entropy, i.e., its localization in momentum space. The possibility or impossibility of overcoming these limitations via TSRR is considered, with respect to both standard and absorption TSRR. (In standard TSRR, refrigeration is achieved at the expense of work input; in absorption TSRR, at the expense of high-temperature heat input.) We then consider the possibility or impossibility of the attainability of absolute zero temperature via configurational-entropy-reduction refrigeration (CSRR). CSRR entails reduction of a refrigerated system's configurational entropy, i.e., its localization in position space, via positional isolation of entities that happen to be in their ground states. Of course, the Second Law of Thermodynamics requires any decrease in entropy of a refrigerated system to be paid for by a compensating greater (in the limit of perfection, equal) increase in eLtropy. Or, in other words, the Second law of Thermodynamics requires any localization in the total momentum-plus-position phase space of a refrigerated system to be paid for by a compensating greater (in the limit of perfection, equal) delocalization in the total momentum-plus-position phase space of the refrigerated system and/or of its surroundings. We also briefly consider energy-reduction refrigeration (ERR), which entails extraction of energy but not entropy from a refrigerated system, and quantum-control refrigeration (QCR). (S not E denotes entropy in TSRR and CSRR, and E denotes energy in ERR, because S is the standard symbol for entropy, and E for energy.) With respect to both TSRR and CSRR, we consider not only the issue of attainability of absolute zero, but also the separate issues, even if absolute zero can be attained, of maintaining it, and of verifying that it has been attained. Purely dynamic – as opposed to thermodynamic – limitations on the quest for absolute

zero under classical versus quantum mechanics are compared and contrasted. Then hot true and cold effective negative Kelvin temperatures are considered. A few fine points concerning the Third Law of Thermodynamics are briefly mentioned in the Appendix.

Keywords: absolute zero, unattainability formulation of the Third Law of Thermodynamics, quantization, energy-time uncertainty principle, negative Kelvin temperatures.

1. Introduction

We consider first the absolute zero of temperature and then negative Kelvin temperatures.

The unattainability formulation of the Third Law of Thermodynamics is briefly reviewed in Sect. 2.1. It puts limitations of the quest for absolute zero, and in its *strongest mode* forbids the attainment of absolute zero by *any method* whatsoever. But typically it is stated principally with respect to thermal-entropy-reduction refrigeration (TSRR). TSRR entails reduction of a refrigerated system's thermal entropy, i.e., its localization in the momentum part of phase space (in momentum space for short). The possibility or impossibility of overcoming these limitations via TSRR is considered, in Sects. 2.2. and 2.3. with respect to standard TSRR, and in Sect. 2.4. with respect to absorption TSRR. (In standard TSRR, refrigeration is achieved at the expense of work input; in absorption TSRR, at the expense of high-temperature heat input.)

In Sect. 3, we consider the possibility or impossibility of the attainability of absolute zero temperature via configurational-entropy-reduction refrigeration (CSRR). CSRR entails reduction of a refrigerated system's configurational entropy, i.e., its localization in the position part of phase space (in position space for short), via positional isolation of entities that happen to be in their ground states. In TSRR, whether standard or absorption, a refrigerated system's thermal energy, as well as its thermal entropy, is reduced. By contrast, in CSRR only its configurational entropy is reduced: since the entities to be positionally isolated are already in their ground states, their thermal energy can*not* be reduced. In Sect. 3, we consider CSRR via positional isolation, by means of weighing or Stern-Gerlach apparatus, of entities that happen to be in their ground states, with reference to a specific one of the quantum-control-refrigeration (QCR) methods investigated in Ref. [1], but we will not employ this or any other QCR method per se.

Refrigeration of a system is also possible via extraction only of energy, but not of entropy, from this system. We dub this type of refrigeration as energy-reduction refrigeration (ERR). In order for energy to be extracted from a system without entropy being extracted from it, the energy must be extracted solely as work and not at all as heat. Simple examples include a one-time perfect (isentropic, reversible) adiabatic expansion of a gas, with energy but not entropy extracted from the gas solely via its doing work on its surroundings during expansion, and the one-time expansion of the photon gas comprising cosmic background

http://dx.doi.org/10.5772/61641

radiation in an ever-expanding Universe (with no steady-state-theory-type "replacement"). In Sect. 2.1. and especially in Sect. 2.5. we will very briefly discuss one-time-expansion ERR, and in Sect. 3. we will very briefly discuss but not employ another type of ERR that is part of a QCR method. (S not E denotes entropy in TSRR and CSRR, and E denotes energy in ERR, because *S* is the standard symbol for entropy, and *E* for energy.)

Of course, the Second Law of Thermodynamics requires any decrease in entropy of a refrigerated system to be paid for by a compensating greater (in the limit of perfection or reversibility, equal) increase in entropy. Or, in other words, the Second Law of Thermodynamics requires any localization in the total momentum-plus-position phase space of a refrigerated system to be paid for by a compensating greater (in the limit of perfection or reversibility, equal) delocalization in the total momentum-plus-position phase space of the refrigerated system and/or of its surroundings. If *all* of the entropy increase and associated waste heat owing to imperfection or irreversibility can be dumped into the surroundings rather than into a refrigerated system, then the refrigerated system will still be cooled to as low a temperature *as if* refrigeration were perfect (reversible), albeit at higher thermodynamic cost. Although perfect (isentropic, reversible) ERR entails zero *net* entropy change of a refrigerated system, within this zero ERR *does* entail a decrease in this system's thermal or momentum-space entropy (localization in momentum space) and an increase in its configurational or position-space entropy (delocalization in position space). If ERR is imperfect (irreversible) then the increase exceeds the decrease and hence the net entropy change is positive. But again if *all* of the entropy increase and associated waste heat owing to imperfection or irreversibility can be dumped into the surroundings rather than into refrigerated system, then the refrigerated system will still be cooled to as low a temperature *as if* ERR were perfect (reversible), albeit at higher thermodynamic cost.

All other things being equal, only if *all* waste heat owing to irreversibility is dumped outside of the system being refrigerated can imperfect (irreversible) refrigeration by *any* method (standard or absorption TSRR, CSRR, ERR, QCR, etc.) attain a temperature as low as that attainable via perfect (reversible) refrigeration, and then only at higher thermodynamic cost than via perfect (reversible) refrigeration. Otherwise, all other things being equal, imperfect (irreversible) refrigeration cannot attain as low a temperature as that attainable via perfect (reversible) refrigeration, even at higher thermodynamic cost than via perfect (reversible) refrigeration.

The Second Law of Thermodynamics forbids a negative change in total entropy, which would correspond to better-than-perfect refrigeration by *any* method (standard, absorption, or other) TSRR, CSRR, ERR, QCR, etc. (In Sect. 3.6, we will give a brief hypothetical consideration of better-than-perfect refrigeration.)

With respect to both TSRR and CSRR, we consider not only the issue of *attainability* of absolute zero, but also the separate issues, *even if* absolute zero can be *attained*, of *maintaining* it, and of *verifying* that it has been attained. The issues of attaining and maintaining absolute zero are considered in both Sects. 2. and 3. The issues of verifiability and *purely* dynamic — as opposed to *thermo*dynamic — limitations on the quest for absolute zero are considered in Sect. 3, because they seem to be more transparently understandable with respect to CSRR, but in Sect. 3. we then relate them also with respect to TSRR. *Purely* dynamic — as opposed to *thermo*dynamic — limitations on the quest for absolute zero under classical versus quantum

mechanics are compared in Sect. 3. Our considerations in Sects. 2. and 3. are general in nature, rather than of specific technical aspects of any particular refrigeration apparatus.

In Sect. 4, we briefly review hot true negative Kelvin temperatures, and then consider cold effective negative Kelvin temperatures. Brief concluding remarks are provided in Sect. 5. A few fine points concerning the third law of thermodynamics are briefly mentioned in the Appendix.

2. The quest for absolute zero via TSRR

2.1. Limits imposed by the Second Law and by the unattainability formulation of the Third Law on TSRR

According to the unattainability formulation of the Third Law of Thermodynamics, the absolute zero of temperature, 0 K, is unattainable in a finite number of finite operations [2–5].[1] But these operations are usually assumed to be TSRR operations [2–5], and most usually *standard* TSRR operations [2–5]. [It might be argued that a one-time infinite operation, for example a one-time infinite adiabatic expansion of a gas, or of the photon gas comprising cosmic background radiation in an ever-expanding Universe (with no steady-state-theory-type "replacement"), *can* via ERR attain 0 K. But (except perhaps for an ever-expanding Universe) a one-time infinite operation is as physically impossible and physically unrealizable as an infinite number of finite operations. Hence we will not employ one–time-expansion ERR in this chapter. (In this Sect. 2.1. and especially in Sect. 2.5. we will very briefly discuss one–time-expansion ERR. In Sect. 3. we will very briefly discuss but not employ another type of ERR that does not require infinite volume (and that is part of a QCR method) for cooling to 0 K, but which still encounters another difficulty with respect to cooling to 0 K.)]

Standard TSRR most typically entails, first, reducing the position-space or configurational entropy of a system to be refrigerated without a compensating increase in its momentum-space or thermal entropy. This first, isothermal, step is *necessary* but *preparatory*, itself *not* yielding a lowering of temperature. An example is isothermal compression of a gas, the heat of compression being expelled to the surroundings, with the surroundings rather than the system to be refrigerated thus suffering the required compensating increase in momentum-space or thermal entropy. The Second Law of Thermodynamics requires that the system's surroundings must suffer a larger (in the limit of perfection, equal) increase in momentum-space or thermal entropy than the decrease in the refrigerated system's position-space or configurational entropy owing to compression of the gas to within a smaller volume. In standard TSRR this is accomplished via heat transfer from the system to be refrigerated to its surroundings [2–5]. Thus momentum-space or thermal entropy is dumped from the system into its surroundings [2–5]. Other examples include isothermal condensation or magnetization, with the heat and thermal entropy thereby released similarly dumped into the surroundings [2–5]. In the second, adiabatic, step, the system to be refrigerated is thermally isolated, so that it can receive no heat from its surroundings. Then, via doing work

[1] (Re: Entries [2] and [3], Refs. [2] and [3]) Contrary to one minor statement on p. 30 of Ref. [3], internal energy *does* have a uniquely-defined zero, in accordance with $E = mc^2$. Reference [3] does not render Ref. [2] obsolete, because Ref. [2] discusses aspects not discussed in Ref. [3], and vice versa.

http://dx.doi.org/10.5772/61641

on its surroundings and/or internally within itself, the refrigerated system trades an increase in its position-space or configurational entropy for a decrease in its momentum-space or thermal entropy. Examples include adiabatic expansion of a gas, wherein work is done on the surroundings, and adiabatic evaporation or demagnetization, wherein at least some of the work is done internally against attractive forces within the refrigerated system itself. The refrigerated system thus follows an adiabat towards a decrease in its temperature. Note that lowering of temperature occurs in this second step, the first step being necessary but preparatory. The *net* result of *both* steps is a decrease in our refrigerated system's momentum-space or thermal entropy but no change in its position-space or configurational entropy — localization in position space in the first step is undone by delocalization in position space in the second step. Thus the refrigerated system's *net* localization is in momentum space but not in position space: hence the designation TSRR. The second, adiabatic, step of standard TSRR, being the temperature-lowering step, may seem to be the more important one. But it would be impossible without the first, isothermal, preparatory step. Note that TSRR requires *heat* to be extracted from a refrigerated system at *some* point in the refrigeration process, even if not at the temperature-lowering step. In the examples of standard TSRR given in this paragraph, this required extraction of heat occurs in the first, isothermal, preparatory step.

Standard TSRR is perfect (reversible) if, in the first step, the decrease in the system's position-space or configurational entropy equals the increase in the surroundings' momentum-space or thermal entropy, and if, in the second step, the decrease in the system's momentum-space or thermal entropy equals the increase in the system's position-space or configurational entropy. Thus standard TSRR is perfect (reversible) if the *total* entropy change is zero for each step considered individually and for both steps combined. TSRR is imperfect (irreversible) if the total entropy change is positive. As per the fourth through sixth paragraphs of Sect. 1. applied specifically to TSRR, all other things being equal, only if *all* waste heat is dumped outside of the system being refrigerated can imperfect (irreversible) TSRR attain a temperature as low as that attainable via perfect (reversible) TSRR, and then only at greater thermodynamic cost than via perfect (reversible) TSRR. Otherwise, all other things being equal, imperfect (irreversible) TSRR cannot attain as low a temperature as that attainable via perfect (reversible) TSRR.

We will consider two types of TSRR. *Standard* TSRR, which we described briefly in the first two paragraphs of this Sect. 2.1, and which we will consider more detail in Sects. 2.2. and 2.3, is executed at the expense of *work* input. In standard TSRR, heat is extracted from a refrigerated system at the expense of *work* input. In Sect. 2.4, we will provide a brief comparison with *absorption* TSRR, which is executed at the expense of *heat* input from a high-temperature reservoir. In absorption TSRR, heat is extracted from a refrigerated system at the expense of *heat* input from a high-temperature reservoir.

To re-emphasize, TSRR, whether standard or absorption, always requires energy to be extracted from a refrigerated system via *heat* at *some* point in the refrigeration process, even if not at the temperature-lowering step. In standard TSRR methods described in the first paragraph of this Sect. 2.1. heat must be extracted from a refrigerated system in the first step to maintain the system isothermal despite compression and/or other localization in position space, even though no heat is extracted from it in the second, adiabatic, temperature-lowering step. Energy may also be extracted from a refrigerated system via work during standard

TSRR. As we will see in Sect. 2.4, in absorption TSRR energy is extracted from a refrigerated system continuously via heat, but never via work.

Because entropy changes become ever smaller as 0 K is approached [2–5], and also because the rate of change of entropy with respect to temperature never becomes infinite [3], the temperature decrease attainable with each successive two-step isothermal-adiabatic standard-TSRR cycle becomes ever smaller [2–5]. Two adiabats can never intercept, and in particular no other adiabat can intercept that corresponding to zero-point entropy and hence to 0 K [2–5]: This is probably the *paramount* reason why, according to the unattainability formulation of the Third Law, of Thermodynamics 0 K cannot be reached in a finite number of finite standard-TSRR operations; the reasons cited in the first sentence of this paragraph probably being more of a supplementary nature [2–5]. [It might be argued that a one-time infinite adiabatic expansion of a gas or of the photon gas comprising cosmic background radiation in an ever-expanding Universe (with no steady-state-theory-type "replacement"), can via ERR attain 0 K, and thus via such infinite expansion its adiabat can intercept that corresponding to zero-point entropy and hence to 0 K. But (except perhaps for an ever-expanding Universe), a one-time infinite expansion is as physically impossible and physically unrealizable as an infinite number of finite operations.]

Beyond these considerations [2–5] concerning limitations on the quest for 0 K in the classical regime, there is of course a vast literature concerning the quest for 0 K, as well as concerning optimizing refrigerator operation, in the quantum regime. Extensive and thorough discussions, reviews, and bibliographies are provided in Refs. [6] and [7].[2] We also cite one specific study [8] (of very many). This study [8] investigates optimization of quantum refrigerator operation via maximization of the product of the time rate of heat extraction from a refrigerated system and the thermodynamic efficiency (coefficient of performance) of the refrigerator, maximization of both simultaneously being impossible [8]. The bottom line based on these sources [6–8] seems to be that even in the quantum regime 0 K cannot be attained with finite resources and in finite time [6–8]. Thus, based on these sources [6–8], quantum refrigeration also seems to be under the governance of the unattainability statement of the Third Law, at least in its strongest mode [6–8]. Yet there is an alternative viewpoint [1].

2.2. The Third Law does *not* require *infinite* work to *attain* $T_C = 0\,\mathrm{K}$ via standard TSRR, but forbids performance of the required *finite* work

The work required to cool any finite sample of matter (and/or of energy such as equilibrium blackbody radiation) maintained within a fixed finite volume V or at constant pressure P from any initial finite fixed, relatively hot ambient temperature T_H to what is generally considered to be the ultimate cold temperature $T_C = 0\,\mathrm{K}$ via *standard* TSRR is *finite* — indeed for typical room-temperature T_H and for typical laboratory-size samples typically *small*. Hence the unattainability formulation of the Third Law of Thermodynamics does *not* forbid *attainment* of 0 K via standard TSRR *by requiring infinite work* for the process. Rather, it forbids *attainment* of 0 K via standard TSRR by forbidding the performance of the required *finite*, typically *small*, amount of work.

[2] (Re: Entry [6], Ref. [6]) The "quest for absolute zero" in the titles of Sects. 2. and 3. of this chapter was borrowed from the titles of Refs. 15 and 36 cited in Ref. [6] of this chapter.

http://dx.doi.org/10.5772/61641

While the coefficient of performance of standard TSRR decreases towards 0 as 0 K is approached, specific heats and heat capacities decrease even more rapidly towards 0 as 0 K is approached. Hence the *finite* — indeed typically *small* — amount of work required to *attain* 0 K via standard TSRR is *not* an issue. Let the cold temperature of a refrigerated system at a given stage of a standard-TSRR process be T_C (T_H assumed fixed). Let dQ_C be the maximum differential increment of heat that the Second Law of Thermodynamics allows to be extracted from this refrigerated system at temperature T_C, with differential increment of heat dQ_H ejected at temperature T_H, at the expense of a given differential increment of work dW. By the First Law of thermodynamics

$$dQ_H = dQ_C + dW. \tag{1}$$

The best possible standard-TSRR operation and hence the highest possible standard-TSRR coefficient of performance COP_{std} allowed by the Second Law of Thermodynamics is in accordance with [9,10]

$$\begin{aligned} dS_{\text{total}} = dS_C + dS_H = \frac{dQ_C}{T_C} - \frac{dQ_H}{T_H} = \frac{dQ_C}{T_C} - \frac{dQ_C + dW}{T_H} = 0 \\ \implies \frac{dW}{T_H} = dQ_C \left(\frac{1}{T_C} - \frac{1}{T_H} \right) = dQ_C \frac{T_H - T_C}{T_C T_H} \\ \implies dW = dQ_C \frac{T_H - T_C}{T_C} \\ \implies COP_{\text{std}} = \frac{dQ_C}{dW} = \frac{T_C}{T_H - T_C} \\ \implies \lim_{T_C \to 0\,\text{K}} COP_{\text{std}} = \frac{T_C}{T_H}. \end{aligned} \tag{2}$$

In the third step of the first line line of Eq. (2) we applied the First Law of Thermodynamics [Eq. (1)]. In the last two lines of Eq. (2) the coefficient of performance COP_{std} for standard TSRR is given in general and then in the limiting case $T_C \to 0\,\text{K}$.

Let Q_C be the heat that must be extracted from our refrigerated system to cool it from T_H to 0 K, $C_X(T_C)$ be the heat capacity given condition X ($X = V$ if constant volume, $X = P$ if constant pressure) of this system at temperature T_C as its temperature T_C is lowered from T_H towards 0 K, and W be the minimum work that the Second Law of Thermodynamics requires to cool it from T_H to 0 K via standard TSRR. Then

$$Q_C = \int_{0\,\text{K}}^{T_H} dQ_C = \int_{0\,\text{K}}^{T_H} C_X(T_C)\, dT_C \tag{3}$$

and

$$\begin{aligned} dW = \frac{dQ_C}{COP_{\text{std}}} = \frac{T_H - T_C}{T_C} dQ_C = \frac{T_H - T_C}{T_C} C_X(T_C)\, dT_C < \frac{T_H}{T_C} dQ_C = \frac{T_H}{T_C} C_X(T_C)\, dT_C \\ \implies W = \int_{0\,\text{K}}^{T_H} dW = \int_{0\,\text{K}}^{T_H} \frac{T_H - T_C}{T_C} C_X(T_C)\, dT_C < T_H \int_{0\,\text{K}}^{T_H} \frac{C_X(T_C)}{T_C} dT_C. \end{aligned} \tag{4}$$

It has been stated [12,13]: "The coefficient of performance becomes progressively smaller as the temperature T_C decreases relative to T_H. And if the temperature T_C approaches zero, the coefficient of performance also approaches zero (assuming T_H fixed). It therefore requires huge amounts of work to extract even trivially small quantities of heat from a system near $T_C = 0\,\mathrm{K}$." But the quantities of heat that must be extracted from *any* laboratory-size system as $T_C \to 0\,\mathrm{K}$ are *less* than trivially small; hence the required amounts of work are *less* than huge. If $C_X(T_C)$ were constant, independent of T_C, then by Eq. (4) the minimum work W required by the second law for cooling *any* finite system to 0 K, i.e., for reducing T_C from T_H to 0 K, would indeed diverge towards ∞ — but even then just barely (logarithmically) — as $T_C \to 0\,\mathrm{K}$. But if $C_X(T_C)$ decreases *at all* — even however slowly — with decreasing T_C, then by Eq. (4) the divergence of W as $T_C \to 0\,\mathrm{K}$ is cured. In fact $C_X(T_C)$ not merely decreases but decreases *rapidly* [14,15] as $T_C \to 0\,\mathrm{K}$; hence the divergence of W as $T_C \to 0\,\mathrm{K}$ is not merely cured but cured by an extremely wide margin. *Any* system constrained within a fixed finite volume, or even within an unfixed but always finite volume for example corresponding to maintenance of constant pressure, *must* obey the two-state model [16] in the limit $T_C \to 0\,\mathrm{K}$ [16], because (as will be discussed in Sect. 3.4.) quantum mechanics *requires* discrete energy levels and *forbids* an energy continuum in any system constrained within a fixed finite volume, or even within an unfixed but always finite volume for example corresponding to maintenance of constant pressure.[3] And for the two-state model [16] the heat capacity decreases nearly exponentially with decreasing temperature in the limit $T_C \to 0\,\mathrm{K}$ [16]. We note that if T_H is not too high it makes little difference whether we put $C_X(T_C) = C_V(T_C)$ or $C_X(T_C) = C_P(T_C)$ in Eqs. (3) and (4) [14,15], because always $T_C \leq T_H$, with $C_P(T_C) - C_V(T_C) \to 0$ and $C_P(T_C)/C_V(T_C) \to 1$ (both from above) as $T_C \to 0\,\mathrm{K}$ [14,15]. Indeed, since solids and liquids are typically only slightly compressible, for solids and liquids $C_P(T_C)$ is typically only marginally larger than $C_V(T_C)$ even at T_C well above 0 K, even as far above 0 K as is consistent with the existence of solids and liquids [17,18].

But as was discussed in Sect. 2.1, in accordance with the unattainability formulation of the third law of thermodynamics [2–5], entropy changes become ever smaller as 0 K is approached, rates of change of entropy with respect to temperature never become infinite [3], and two adiabats can never intercept [2–5]. Hence the temperature decrease attainable with each successive two-step isothermal-adiabatic TSRR cycle becomes ever smaller [2–5]. Especially, two adiabats can never intercept, and in particular no other adiabat can intercept that corresponding to zero-point entropy and hence to 0 K [2–5]. Thus the unattainability formulation of the third law forbids the attainment of 0 K via standard TSRR *not* by requiring an *infinite* amount of work to cool a finite sample of matter (and/or energy) within a finite volume to 0 K, but rather by forbidding the *finite* — typically *small* — required amount of work from being performed via any standard-TSRR process [2–5]. While COP_{std} as per Eq. (2) is the theoretical maximum and the work W as per Eq. (4) is the theoretical minimum allowed by the Second Law of Thermodynamics, for well-designed real-world standard TSRR systems the actual COP_{std} is not more than a few times smaller and the actual required work

[3] (Re: Entry [16], Ref. [3]) In the version of the two-state model presented in Sects. 15-3 and 16-2 of Ref. [3], both states are nondegenerate. But the nearly exponential decrease of heat capacity with decreasing temperature as $T \to 0\,\mathrm{K}$ still obtains irrespective of any (finite) degeneracy of one or both states. Degeneracy of the ground state does not affect the heat capacity at all; g-fold degeneracy of the excited state multiplies g-fold the heat capacity as compared to that given a nondegenerate excited state as per Sect. 15-3 of Ref. [3].

http://dx.doi.org/10.5772/61641

is not more than a few times larger these theoretical limits, and this small numerical factor in no way contravenes our result.

2.3. Can the difficulty of infinite *power* required to *maintain* $T_C = 0\,\mathrm{K}$ be overcome?

The work W required to *attain* $T_C = 0\,\mathrm{K}$ as derived in Sect. 2.2. is that required *only* to *extract* all of the *internal* thermal energy *out of* a system to be refrigerated via standard TSRR, thus cooling it to $T_C = 0\,\mathrm{K}$. But this *un*realistically assumes that *strictly zero external* thermal energy flows *into* this system in the meantime. Thus we must consider *not only* the work W required to *extract* all of the *internal* thermal energy *out of* a system to be refrigerated, thus cooling it to $T_C = 0\,\mathrm{K}$, i.e., to *attain* $T_C = 0\,\mathrm{K}$. We must *also* consider the power $P = dW'/dt$ and work $W' = \int P dt$ required to overcome the flow of *external* thermal energy, i.e., heat flow dQ_C/dt, *into* our refrigerated system, requisite to *maintain* it at $T_C = 0\,\mathrm{K}$. (Note: Time t should not be confused with temperature T.) Heat transfer into our refrigerated system, indeed heat transfer in general, occurs via three processes: conduction, radiation, and convection [19,20]. Heat transfer dQ_C/dt into our refrigerated system via conduction is proportional to $T_H - T_C$ and via radiation to $T_H^4 - T_C^4$ [19,20]. While convection is a complex phenomenon, which may be either natural or forced, for simplicity and for argument's sake let us accept the most usual result [21], according to which heat transfer via natural convection is proportional to $(T_H - T_C)^{5/4}$ [21]. If we must have convection, we prefer natural convection to forced convection, because the former transfers heat less efficiently. Thus heat transfer dQ_C/dt via conduction is $a(T_H - T_C)$, via radiation $b\left(T_H^4 - T_C^4\right)$, and via natural convection $c(T_H - T_C)^{5/4}$ [19,20]; the prefactors a, b, and c corresponding to conductive, radiative, and natural-convective heat transfer, respectively [19,20]. These three prefactors for a given system to be refrigerated are determined by its geometry (size, shape, surface area, etc.), by the type of insulation, usually at least to some extent by T_H and/or T_C, and by any other relevant properties [19,20]. Thus, applying the result for COP_{std} from Eq. (2),

$$\begin{aligned}
&\frac{dQ_C}{dt} = a(T_H - T_C) + b\left(T_H^4 - T_C^4\right) + c(T_H - T_C)^{5/4} \\
&\implies P = \frac{dW'}{dt} = \frac{dQ_C/dt}{COP_{\mathrm{std}}} = \frac{T_H - T_C}{T_C}\frac{dQ_C}{dt} = \frac{T_H - T_C}{T_C}\left[a(T_H - T_C) + b\left(T_H^4 - T_C^4\right) + c(T_H - T_C)^{5/4}\right] \\
&\implies \lim_{T_C \to 0\,\mathrm{K}} \frac{dQ_C}{dt} = aT_H + bT_H^4 + cT_H^{5/4} \\
&\implies \lim_{T_C \to 0\,\mathrm{K}} P = \lim_{T_C \to 0\,\mathrm{K}} \frac{dW'}{dt} = \lim_{T_C \to 0\,\mathrm{K}} \frac{dQ_C/dt}{COP_{\mathrm{std}}} = \lim_{T_C \to 0\,\mathrm{K}} \frac{T_H - T_C}{T_C}\frac{dQ_C}{dt} \\
&= \lim_{T_C \to 0\,\mathrm{K}} \frac{T_H}{T_C}\left(aT_H + bT_H^4 + cT_H^{5/4}\right) = \lim_{T_C \to 0\,\mathrm{K}} \frac{aT_H^2 + bT_H^5 + cT_H^{9/4}}{T_C}.
\end{aligned} \tag{5}$$

This expression diverges towards ∞ as $T_C \to 0\,\mathrm{K}$ unless a, b, and c decrease with decreasing T_C at least as rapidly as T_C itself, with $a = b = c = 0$ at least at $T_C = 0\,\mathrm{K}$. But given that a, b, and c in general depend on T_H as well as on T_C and that T_H is fixed, such a functional dependency of a, b, and c on T_C seems unlikely. But perhaps we should not *a priori* rule it out as impossible.

The decrease, indeed the typically *rapid* decrease, of $C_X(T_C)$ as $T_C \to 0\,\mathrm{K}$ [14,15] allows *all* of a finite refrigerated system's *internal* thermal energy to be pumped *out* of it thus cooling it to $T_C = 0\,\mathrm{K}$ via standard TSRR with the expenditure of a finite (typically small) amount of work. The vanishing of $C_X(T_C)$ as $T_C \to 0\,\mathrm{K}$ within a refrigerated system more than compensates for the vanishing of COP_{std} in Eq. (4) as $T_C \to 0\,\mathrm{K}$. But the vanishing of $C_X(T_C)$ as $T_C \to 0\,\mathrm{K}$ within a refrigerated system does *not* help insofar as overcoming the flow of *external* thermal energy, i.e., heat flow, from surroundings at ambient temperature T_H *into* a refrigerated system, is concerned. Thus in Eq. (5) the vanishing of COP_{std} as $T_C \to 0\,\mathrm{K}$ is *not* compensated for. Hence *even if* $T_C = 0\,\mathrm{K}$ is *attained* infinite power is required to *maintain* it — *unless* a, b, and c decrease with decreasing T_C at least as rapidly as T_C itself, with $a = b = c = 0$ at least at $T_C = 0\,\mathrm{K}$. Can this "*unless*" — implying that insulation must become *perfect* [23] as $T_C \to 0\,\mathrm{K}$ — be realized? Again, given that a, b, and c in general depend on T_H as well as on T_C and that T_H is fixed, such a functional dependency of a, b, and c on T_C, it seems unlikely, but perhaps we should not *a priori* rule it out as impossible. But *even if* it is impossible, $T_C = 0\,\mathrm{K}$ may be not merely *attainable*, but also *maintainable*, *but only for an instant, or at most for a finite number of instants*.

Note that $COP_{\mathrm{std}} = 0$ obtains *only* at *exactly* the *point* value $T_C = 0\,\mathrm{K}$ [9,10]. This leaves open the possibility that even if a, b, and c do *not* decrease with decreasing T_C at least as rapidly as T_C itself and remain finite and positive at $T_C = 0\,\mathrm{K}$, i.e., even if insulation does *not* become perfect as $T_C \to 0\,\mathrm{K}$, $T_C = 0\,\mathrm{K}$ could be *attained* and then *maintained for an instant*, because P need be infinite for only an infinitesimally short time so $W' = \int P dt$ could still be finite. But if even given this *im*perfection $T_C = 0\,\mathrm{K}$ could thus be attained *even for an instant*, then it could be likewise re-attained for any arbitrarily large (but finite) number $\mathfrak{N}$ of additional instants, since $\mathfrak{N} \int P dt$ would then still be finite. Any finite number of infinitesimally short time intervals still sum to an infinitesimally short time interval. Let the refrigeration process begin at time t_0 and be completed at time t_1. Let T_H be fixed, and for simplicity and for argument's sake let

$$T_C(t) = T_H\left(1 - \frac{t - t_0}{t_1 - t_0}\right)^{\gamma}, \tag{6}$$

where γ is a fixed positive real number. Then, for simplicity and for argument's sake, let us consider only the part of the refrigeration process at $T_C \ll T_H$. Within this part of the refrigeration process, T_C has little room to decrease towards $T_C = 0\,\mathrm{K}$, so since T_H is fixed letting a, b, and c be constants independent of T_C may be a good approximation. Thus we have, applying the last two lines of Eq. (5),

$$\begin{aligned}
W' &= \int_{t_0}^{t_1} P dt = \left(aT_H^2 + bT_H^5 + cT_H^{9/4}\right) \int_{t_0}^{t_1} \frac{dt}{T_C(t)} \\
&= \left(aT_H^2 + bT_H^5 + cT_H^{9/4}\right) \int_{t_0}^{t_1} \frac{dt}{T_H\left(1 - \frac{t-t_0}{t_1-t_0}\right)^{\gamma}} = \left(aT_H + bT_H^4 + cT_H^{5/4}\right) \int_{t_0}^{t_1} \frac{dt}{\left(1 - \frac{t-t_0}{t_1-t_0}\right)^{\gamma}} \\
&= \left(aT_H + bT_H^4 + cT_H^{5/4}\right) \int_{t_0}^{t_1} \frac{dt}{\left[\frac{t_1-t_0-(t-t_0)}{t_1-t_0}\right]^{\gamma}} = \left(aT_H + bT_H^4 + cT_H^{5/4}\right)(t_1 - t_0)^{\gamma} \int_{t_0}^{t_1} \frac{dt}{(t_1 - t)^{\gamma}}
\end{aligned}$$

http://dx.doi.org/10.5772/61641

$$\begin{aligned}
&\implies \text{if } 0<\gamma<1 \text{ then } W' = \left(aT_H + bT_H^4 + cT_H^{5/4}\right)(t_1-t_0)^{\gamma}\,\frac{(t_1-t_0)^{1-\gamma}}{1-\gamma} = \\
&= \left(aT_H + bT_H^4 + cT_H^{5/4}\right)\frac{t_1-t_0}{1-\gamma} \\
&\implies W' \text{ is finite if } 0<\gamma<1 \\
&\implies W_{\text{total}} = W + W' \text{ is finite if } 0<\gamma<1. \qquad (7)
\end{aligned}$$

In the last step of Eq. (7) we applied the finiteness of our result for W as per Eq. (4) and the associated discussions. Thus it seems that the difficulty of infinite power and infinite work required to *maintain* $T_C = 0\,\text{K}$ can, at least to this very limited extent, be overcome.

W' and hence $W_{\text{total}} = W + W'$ can remain finite if $T_C = 0\,\text{K}$ is to be maintained for *finitely longer than* an instant or finite number of instants given given fixed finite $T_H > 0\,\text{K}$ only if a, b, and c decrease with decreasing T_C at least as rapidly as T_C itself, with $a = b = c = 0$ at least at $T_C = 0\,\text{K}$. And this in the face of a, b, and c in general depending on T_H as well as on T_C, with T_H being fixed. Thus we require *perfect* [23] — not merely good — insulation at $T_C = 0\,\text{K}$, yet also with fixed finite $T_H > 0\,\text{K}$. And perfect insulation — $a = b = c = 0$ — is hard to come by. Hard, but perhaps not impossible. Again, for simplicity and for argument's sake, let us consider only the part of the refrigeration process at $T_C \ll T_H$. By surrounding our refrigerated system including its insulation with a vacuum, and with the insulation being comprised entirely of solids (not fluids: liquids or gases), we can indeed achieve $c = 0$ — solids certainly exist at finite $T_H > 0\,\text{K}$. Convection (whether natural or forced) occurs only in fluids (gases and liquids), and is nonexistent in solids or in a vacuum. But even though we have thus achieved $c = 0$, we must still achieve $a = 0$ and $b = 0$. Superinsulators — perfect (not merely good) — insulators with respect to electricity have recently been discovered [23], with the superinsulating state existing at temperatures up to $T_{SI,\text{elec,max}}$ finitely greater than $0\,\text{K}$ [23]. (Reference [23] provides a thorough and excellent review, as well as an extensive bibliography.) So perhaps we should not *a priori* rule out superinsulators with respect to heat, with the superinsulating state existing at $0\,\text{K} \leq T_H \leq T_{SI,\text{heat,max}}$ [23]. Even if superinsulation with respect to heat exists, we do not know if $T_{SI,\text{heat,max}} = T_{SI,\text{elec,max}}$. But all we require is that $0\,\text{K} < T_H < T_{SI,\text{heat,max}}$ [23]. If superinsulation with respect to heat exists, then $a = 0$ can obtain at finite $T_H > 0\,\text{K}$. [Superinsulation should not be confused with the typical exponential improvement of ordinary insulation with decreasing temperature. The latter obtains, for example, if conduction of heat and electricity is via electrons thermally promoted from the valence band to the conduction band with the two bands separated by a fixed finite energy gap ΔE. Then the probability of such promotion per attempt to jump the gap decreases exponentially with decreasing T_H in accordance with the Boltzmann factor $e^{-\Delta E/kT_H}$ (k is Boltzmann's constant) and hence is very small for low T_H. But it does not vanish *perfectly* except at $T_H = 0\,\text{K}$ and hence ordinary insulation remains *im*perfect at any finite $T_H > 0\,\text{K}$.] If furthermore the vacuum surrounding our refrigerated system with its superinsulating shield with respect to heat is permeated by equilibrium blackbody radiation at fixed finite T_H below the upper temperature limit $T_{SI,\text{heat,max}}$ of the superinsulating state with respect to heat, then this radiation will not destroy the superinsulating state. (Indeed a vacuum *must* be permeated by equilibrium blackbody radiation at any temperature finitely greater than $0\,\text{K}$.) If the superinsulator is opaque to this equilibrium blackbody radiation,

with the radiation being thermalized in its outer layer to internal energy at $T_H < T_{SI,\text{heat,max}}$, or scattered or reflected away, then $b = 0$. A nonopaque superinsulator can be shielded by an opaque material, with the radiation being thermalized in the outer layer of this opaque material to internal energy at $T_H < T_{SI,\text{max}}$, or scattered or reflected away, so that $b = 0$. Of course for $b = 0$ *exactly* the opacity must be not merely good but *perfect*. While this perfection may be impossible to achieve *exactly*, it can be achieved *for all practical purposes*. Typically the fraction of incident radiation not thermalized as internal energy within an opaque material, or not scattered or reflected away, decreases exponentially increasing thickness of an opaque material. An incident photon has a probability of $e^{-\mathcal{N}}$ of penetrating through a thickness of $\mathcal{N}$ or more *e*-folding lengths without being thermalized as internal energy within an opaque material, or being scattered or reflected away. If, say, $\mathcal{N} \gtrsim 1000$, then for all practical purposes we can rest assured that not even 1 photon will get through during the time required for any refrigeration experiment and hence that, even if not exactly then for all practical purposes, $b = 0$.

Thus at least *prima facie* it seems that there seems to be no difficulty in principle in achieving $c = 0$, and even if not perfectly then for all practical purposes also $b = 0$. The main concern is whether or not $a = 0$ is achievable, namely whether or not superinsulation exists with respect to heat as it does with respect to electricity [23]. Probably the best that we can do at this point is to admit that we do not know; that this is an open question [23].

2.4. A brief comparison with absorption TSRR

The discussions in Sects. 2.2. and 2.3. presuppose *standard* TSRR, which operates as a heat engine in reverse. In heat engine operation heat flows from a hot reservoir via the engine into a cold reservoir; within the limit imposed by the Second Law of Thermodynamics, the engine can convert part of this heat flow into work output. Standard TSRR operates as a heat engine in reverse, with work input driving heat flow from a cold reservoir into a hot one, also within the limit imposed by the Second Law of Thermodynamics.

However, there is one other commonly-employed type of TSRR that we wish to consider — *absorption* TSRR [24]. While absorption TSRR is not employed in practice to reach cryogenic temperatures, let alone to approach 0 K, it may be of interest to consider it even if only in principle. Absorption TSRR requires zero work input [24]. Instead, heat Q_H is supplied to the refrigeration apparatus from a hot reservoir at temperature T_H, heat Q_C is extracted by the refrigeration apparatus from a refrigerated system at cold temperature T_C, and heat Q_I is ejected from the refrigeration apparatus at intermediate temperature T_I ($T_H > T_I > T_C$) [24]. By the First Law of Thermodynamics

$$dQ_I = dQ_C + dQ_H. \tag{8}$$

Let the cold temperature of a refrigerated system at a given stage of an absorption-TSRR process be T_C (T_I and T_H assumed fixed). Let dQ_C be the maximum differential increment of heat that the Second Law of Thermodynamics allows to be extracted from this refrigerated system at temperature T_C at the expense of a given differential increment of heat input dQ_H at temperature T_H, with differential increment of heat $dQ_I = dQ_C + dQ_H$ ejected at temperature T_I. The best possible absorption-TSRR operation and hence the highest

http://dx.doi.org/10.5772/61641

possible absorption-TSRR coefficient of performance COP_{abs} allowed by the Second Law of Thermodynamics is in accordance with

$$\begin{aligned}
&dS_{\text{total}} = dS_I + dS_C + dS_H = \frac{dQ_I}{T_I} - \frac{dQ_C}{T_C} - \frac{dQ_H}{T_H} = 0 \\
&\implies \frac{dQ_C + dQ_H}{T_I} - \frac{dQ_C}{T_C} - \frac{dQ_H}{T_H} = \frac{dQ_C}{T_I} + \frac{dQ_H}{T_I} - \frac{dQ_C}{T_C} - \frac{dQ_H}{T_H} = 0 \\
&\implies dQ_C \left(\frac{1}{T_C} - \frac{1}{T_I}\right) = dQ_H \left(\frac{1}{T_I} - \frac{1}{T_H}\right) \\
&\implies dQ_C \frac{T_I - T_C}{T_I T_C} = dQ_H \frac{T_H - T_I}{T_I T_H} \\
&\implies dQ_C \frac{T_I - T_C}{T_C} = dQ_H \frac{T_H - T_I}{T_H} \\
&\implies COP_{\text{abs}} = \frac{dQ_C}{dQ_H} = \frac{T_C\,(T_H - T_I)}{T_H\,(T_I - T_C)} \\
&\implies \lim_{T_C \to 0\,\text{K}} COP_{\text{abs}} = \frac{T_C\,(T_H - T_I)}{T_I T_H}.
\end{aligned} \tag{9}$$

In the second line of Eq. (9) we applied the First Law of Thermodynamics [Eq. (8)]. In the last two lines of Eq. (9) the coefficient of performance COP_{abs} for absorption TSRR is given in general and then in the limiting case $T_C \to 0\,\text{K}$.

Now let us compare COP_{abs} with COP_{std}. By comparing Eqs. (2) and (9), we obtain

$$\begin{aligned}
&\frac{COP_{\text{abs}}}{COP_{\text{std}}} = \frac{\frac{T_C(T_H - T_I)}{T_H(T_I - T_C)}}{\frac{T_C}{T_H - T_C}} = \frac{(T_H - T_I)\,(T_H - T_C)}{T_H\,(T_I - T_C)} = \left(1 - \frac{T_C}{T_H}\right)\frac{T_H - T_I}{T_I - T_C} = \left(1 - \frac{T_I}{T_H}\right)\frac{T_H - T_C}{T_I - T_C} \\
&\implies \lim_{T_C \to 0\,\text{K}} \frac{COP_{\text{abs}}}{COP_{\text{std}}} = \frac{T_H - T_I}{T_I} = \frac{T_H}{T_I} - 1.
\end{aligned} \tag{10}$$

Thus

$$\begin{aligned}
&COP_{\text{abs}} > COP_{\text{std}} \text{ if } (T_H - T_I)\,(T_H - T_C) > T_H\,(T_I - T_C) \\
&\implies T_H^2 - T_I T_H - T_C T_H + T_C T_I > T_I T_H - T_C T_H \\
&\implies T_H^2 - 2T_I T_H + T_C T_I > 0 \\
&\implies T_I\,(T_C - 2T_H) > -T_H^2 \\
&\implies T_I\,(2T_H - T_C) < T_H^2 \\
&\implies T_I < \frac{T_H^2}{2T_H - T_C} \\
&\implies T_I < \frac{T_H}{2 - \frac{T_C}{T_H}} \\
&\implies T_I < \frac{T_H}{2} \text{ if } T_C \ll T_H.
\end{aligned} \tag{11}$$

Thus it may be of interest to consider absorption TSRR, even if only in principle, because [24]: (a) It is thermodynamically less costly to supply a *given* quantity of energy input as *heat*, which is sufficient for absorption TSRR, than as *work*, which is required for standard TSRR. (b) If Inequality (11) is fulfilled, then absorption TSRR requires a *smaller* quantity of energy input as *heat* than standard TSRR does as *work*. (c) Some absorption-TSRR systems, notably the Munters/von-Platen system [24] and the Einstein/Szilárd system [25] (which however at least in their original forms cannot attain cryogenic temperatures, let alone approach 0 K), have no moving parts, which minimizes waste of negentropy and free energy via friction while maximizing reliability; also, they operate essentially silently, thus wasting essentially no negentropy and free energy as sound. By Inequality (11), the upper limit of T_I consistent with $COP_{\text{abs}} > COP_{\text{std}}$, i.e., with absorption TSRR requiring less heat input than standard TSRR does work input for given T_H and T_C, never falls below $T_H/2$ even in the limit $T_C \to 0\,\text{K}$. Hence even after T_C has been reduced sufficiently that the last lines of Eqs. (2), (9), and (10) and Inequality (11) are applicable, if $T_I < T_H/2$ then the quantities of Q_H, dQ_H/dt, and $Q'_H = \int \frac{dQ_H}{dt} dt$ required for absorption TSRR are *smaller* than those of W, P, and W', respectively, required in accordance with Eqs. (4), (5), and (7), respectively, for standard TSRR — besides being thermodynamically less costly *per given quantity*. Thus it seems that perhaps we should not *a priori* rule out that approaching or even attaining $T_C = 0\,\text{K}$ may be easier, at least in principle even if not in practice, via absorption TSRR, or perhaps via some variant or modification thereof, than via standard TSRR.

Nevertheless, for T_I finitely greater than 0 K (of course $T_I > T_C$) the advantage of absorption TSRR over standard TSRR is finite. Hence we should restate the first paragraph of Sect. 2.2. with respect to absorption TSRR: The heat input Q_H required to cool any finite sample of matter (and/or of energy such as equilibrium blackbody radiation) maintained within a fixed finite volume V or at fixed finite pressure P from any initial finite fixed, relatively hot ambient temperature T_H to $T_C = 0\,\text{K}$ via absorption TSRR is *finite* — indeed for typical room-temperature T_H and for typical laboratory-size samples typically *small*. Hence the unattainability formulation of the Third Law of Thermodynamics does *not* forbid attainment of 0 K via absorption TSRR *by requiring infinite* Q_H for the process. Rather, it forbids *attainment* of 0 K via absorption TSRR by forbidding the utilization of the required *finite*, typically *small*, Q_H. While COP_{abs} as per Eq. (9) is the theoretical maximum allowed by the Second Law of Thermodynamics, for well-designed real-world absorption TSRR systems the actual COP_{abs} is not more than a few times smaller this theoretical maximum, and this small numerical factor in no way contravenes our result.

But *even if* $T_C = 0\,\text{K}$ *could* be precisely *attained*, whether via standard, absorption, or other TSRR, the question of *maintaining* $T_C = 0\,\text{K}$ discussed in Sect. 2.3. is still open. *Even if* $T_C = 0\,\text{K}$ *could* be precisely *attained*, whether via standard, absorption, or other TSRR, whether or not it is *maintainable* for finitely longer than the infinitesimally short time allowed in accordance with Eqs. (6) and (7) and the associated discussions is still open. But least in principle even if not in practice *if* $T_C = 0\,\text{K}$ can be *attained*, *then maintaining* $T_C = 0\,\text{K}$, whether this is possible only for infinitesimally short time or for finite time, may be more easily achievable via absorption TSRR than via standard TSRR.

2.5. Brief remarks concerning one–time-expansion ERR

One-time-expansion ERR can, at least in principle, be achieved via a sample of gas at ambient pressure at Earth's surface (or the surface of any other planet with an

http://dx.doi.org/10.5772/61641

atmosphere) being transported to a vacuum (either to a vacuum chamber on the planet or to the vacuum of space), and there being allowed to expand adiabatically. For a given expansion ratio, maximum cooling is attained via a perfect (reversible) adiabatic expansion, wherein the decrease in thermal (momentum-space) entropy exactly offsets the increase in configurational (position-space) entropy owing to expansion. But even an irreversible adiabatic expansion is ERR, because even in an irreversible adiabatic expansion only energy and not entropy is extracted from the gas. But in an irreversible adiabatic expansion some thermal (momentum-space) entropy is created within the gas, so that the decrease in thermal (momentum-space) entropy only partially offsets the increase in configurational (position-space) entropy owing to expansion, and hence refrigeration is less efficient, with less cooling per given expansion ratio, than in the perfect (reversible) case. It might be argued that no vacuum that the gas expands into is perfect and hence its expansion cannot continue indefinitely. But in an ever-expanding Universe (with no steady-state-theory-type "replacement") the surrounding vacuum becomes ever more perfect. But (except perhaps for an ever-expanding Universe), a one-time infinite expansion is as physically impossible and physically unrealizable as an infinite number of finite operations.

It might also be argued that a one-time infinite operation, for example a one-time infinite adiabatic expansion of a gas after infinite time in an ever-expanding Universe (with no steady-state-theory-type "replacement"), or of the photon gas comprising cosmic background radiation after infinite time in an ever-expanding Universe (with no steady-state-theory-type "replacement"), *can* via ERR attain 0 K. But this requires *infinite* resources — *infinite* volume and *infinite* time. Hence it does not contravene the conclusion that absolute zero 0 K can*not* be attained with *finite* resources — not only classically but even in the quantum regime [6–8] that we will consider in Sect. 3. Hence we do not employ one–time-expansion ERR in this chapter. [In Sect. 3. we will very briefly discuss but not employ another type of ERR that does not require infinite volume (and that is part of a QCR method) for cooling to 0 K, but which still encounters another difficulty with respect to cooling to 0 K.]

We note that, not unlike an irreversible adiabatic expansion of a gas, a polytropic expansion thereof intermediate between adiabatic and isothermal can achieve refrigeration, albeit less efficiently, with less cooling per given expansion ratio, than a perfect (reversible) adiabatic one. A polytropic expansion intermediate between adiabatic and isothermal can be construed as ERR, because only energy and not entropy is *extracted from* the gas. Thermal (momentum-space) entropy is *imported into* the gas during a polytropic expansion, so that the decrease in thermal (momentum-space) entropy only partially offsets the increase in configurational (position-space) entropy owing to expansion, and hence refrigeration is less efficient than in the perfect (reversible) adiabatic case. The only difference between an irreversible adiabatic expansion and a polytropic one intermediate between adiabatic and isothermal is that thermal (momentum-space) entropy is generated within the expanding gas in the former case and imported into it in the latter. [Of course, additional irreversibilities can result in thermal (momentum-space) entropy being generated within the expanding gas during a polytropic expansion intermediate between adiabatic and isothermal, thus rendering refrigeration still less efficient.]

Perfect (reversible) one-time-expansion adiabatic ERR of our gas, or even imperfect (irreversible) adiabatic or even polytropic (intermediate between adiabatic and isothermal)

ERR thereof, yields rather than costs work. Of course, this does not count the work that it costs to evacuate its vacuum chamber or to transport it to the vacuum of space.

Of course, except perhaps for the cooling of or in an ever-expanding Universe (with no steady-state-theory-type "replacement"), the difficulties of *maintaining* cold as opposed to merely *attaining* it apply with respect to one–time-expansion ERR as with respect to standard and absorption TSRR. These difficulties also apply with respect to CSRR, QCR, and another type of ERR that is part of a QCR method, all to be considered in Sect. 3.

3. The quest for absolute zero via configurational-entropy-reduction refrigeration (CSRR)

In Sect. 3. we consider the quest for absolute zero, $T_C = 0\,\mathrm{K}$, via configurational-entropy-reduction refrigeration (CSRR), which localizes a refrigerated system in the position part of phase space (in position space for short), as opposed to thermal-entropy-reduction refrigeration (TSRR), which localizes it in the momentum part of phase space (in momentum space for short). Standard TSRR requires extraction of energy from a refrigerated system via *heat* during at least *some* step of the refrigeration process. It may also entail extraction of energy from a refrigerated system via work (recall Sect. .2.1. In absorption TSRR energy is extracted from a refrigerated system continuously via heat, but never via work (recall Sect. 2.4.). By contrast, CSRR entails *no* extraction of energy from a refrigerated system *either via heat or via work*. CSRR requires finite work input to *attain* $T_C = 0\,\mathrm{K}$, even if this work input is employed differently than the work input in standard TSRR as per Sect. 2.2, or than high-temperature heat input in absorption TSRR as per Sect. 2.4. CSRR shares with TSRR the difficulties of *maintaining* $T_C = 0\,\mathrm{K}$ as per Sect. 2.3. and the last paragraphs of Sects. 2.4. and 2.5. But in Sect. 3. let us focus mainly on prospects for and limitations on the quest for *attaining* $T_C = 0\,\mathrm{K}$ via CSRR, comparing these prospects and limitations with those via TSRR. We postpone remarking on the difficulties of *maintaining* $T_C = 0\,\mathrm{K}$ via CSRR until the last two paragraphs of Sect. 3.5.

3.1. Questioning the unattainability formulation of the Third Law of Thermodynamics in toto

The unattainability formulation of the Third Law of Thermodynamics *in toto* — not merely any particular limit(s) imposed thereby — has been questioned [1]. Above all, the question of the attainability of 0 K in a finite number of finite operations (perhaps even in one) by *any method* whatsoever, and hence the status of the unattainability formulation of the Third Law of Thermodynamics in its *strongest* mode, according to which this is impossible, remains open [1,26–28]. Even so, the question of whether or not 0 K is attainable by *any method* whatsoever is sometimes stated to be only of academic interest [27], and it is also sometimes stated that there may be "profound problems [22]" concerning attaining "absolute thermal isolation [22]," i.e., perfect insulation [23], and that infinitely precise measurements [22] may be required to *perfectly verify* [22] that *precisely* 0 K has actually been attained [22].

Yet it has been shown that 0 K may be attainable in a finite number of finite operations (perhaps even in one) via quantum-control-refrigeration (QCR) methods, specifically, employing quantum coherence [1]. This challenges the *strongest-mode* unattainability

http://dx.doi.org/10.5772/61641

formulation of the Third Law of Thermodynamics, which forbids the attainment of 0 K by *any method* whatsoever [1].

In Sect. 3.2, we will first consider CSRR via positional isolation by means of weighing of entities that happen to be in the ground state. Perhaps in principle, even if not in practice, at least *prima facie* this seems to be the simplest possible method of CSRR. So perhaps it may elucidate at least some of the problems of attaining 0 K, and if 0 K can be attained of *verifying* [22] that 0 K has been attained, more easily than the more technically advanced and more practical QCR methods discussed in Ref. [1], which are much more amenable to realization using currently-available technology [1]. We then consider CSRR via positional isolation, by means of a Stern-Gerlach apparatus, of entities that happen to be in the ground state. Our consideration of CSRR via a Stern-Gerlach apparatus will be with reference to a specific one of the QCR methods [1], but we will not employ this or any other QCR method per se. In this regard, in the sixth paragraph of Sect. 3.2. we will briefly discuss but not employ another type of ERR than that discussed in Sect. 2.1 and especially in Sect. 2.5, which entails reduction of a refrigerated system's *non*thermal energy but not of its entropy.

Thus irrespective of the status of TSRR with respect to the unattainability formulation of the Third Law of Thermodynamics, there also exist CSRR methods, which we will consider in this Sect. 3. Even if, as will turn out to at least apparently be the case, even CSRR methods are limited by the strongest-mode unattainability formulation of the Third Law of Thermodynamics, they at least seem to be closer to breaking through this limit than TSRR methods. The ultimate limitation that the unattainability formulation of the Third Law of Thermodynamics can wield in its strongest mode seems to be *purely* dynamic as opposed to *thermo*dynamic — the energy-time uncertainty principle. Thus *exact* attainment of 0 K may be protected against *any* type of refrigeration: TSRR, CSRR, ERR, QCR, or otherwise (or any combination thereof). But the only slightly less ambitious goal of attainment of 0 K *for all practical purposes* seems to be within reach.

3.2. Absolute zero via CSRR (for example isolation in position space by weighing or by Stern-Gerlach apparatus)?

Consider System A comprised of N identical harmonic oscillators, in thermodynamic equilibrium with a heat reservoir at temperature T, an average $\langle n \rangle$ of which are in the ground state. (Averaging is denoted by enclosure within angular brackets.) Let ΔE be the gap between adjacent energy states of any given oscillator. Let T be low enough so that the probability of even one of the harmonic oscillators being in its second or higher excited states is negligible. In accordance with the Boltzmann distribution, the probability $\mathbb{P}_{A1}$ of any given System-A oscillator being in its first excited state is $e^{-\Delta E/kT}$ times the probability $\mathbb{P}_{A0}$ of being in its ground state. (Note: Probability $\mathbb{P}$ should not be confused with power P.) Hence, normalizing yields $\mathbb{P}_{A0} \doteq \langle n \rangle / N = 1/\left(1+e^{-\Delta E/kT}\right)$ and $\mathbb{P}_{A1} \doteq (N - \langle n \rangle)/N = 1 - (\langle n \rangle / N) = 1 - \left[1/\left(1+e^{-\Delta E/kT}\right)\right] = e^{-\Delta E/kT}/\left(1+e^{-\Delta E/kT}\right) \doteq e^{-\Delta E/kT}$. [The dot-equal sign ($\doteq$) means "very nearly equal to."] Of course, T being small enough so that the probability of even one of the harmonic oscillators being in its second or higher excited states is negligible typically implies that $\mathbb{P}_{A1} \ll 1$. But if N is moderately but not excessively large this can obtain consistently with $N\mathbb{P}_{A1} = N - \langle n \rangle$, the average number of oscillators in the first excited state, exceeding unity.

An oscillator in the first excited state has a mass exceeding that of one in the ground state by $\Delta E/c^2$, and, letting g be the local acceleration due to gravity, a weight exceeding that of one in the ground state by $g\Delta E/c^2$. (From now on unless otherwise noted we take c to be the speed of light in vacuum, not the prefactor defined in Sect. 2.3.) Thus (in principle!) the oscillators in the ground state in our original System A at temperature T can be positionally isolated by weighing from those in the first excited state therein — creating in only one operation (albeit consisting of n weighing steps) Subsystem B comprised of n oscillators ($n \leq N$), *all* of which are in the ground state. *Prima facie* it seems that Subsystem B is therefore indeed at the absolute zero of temperature, 0 K. Moreover such positional isolation can in principle be executed via employment only of work interactions and hence with zero heat transfer, either into our ground-state-only Subsystem B or otherwise.

The required work is modest. The entropy — more correctly, *neg*entropy — cost of isolating the first of the n ground-state oscillators is $\Delta S_{\text{isol},1} = k \ln \frac{N}{n}$. The negentropy cost of isolating the second of the n ground-state oscillators is $\Delta S_{\text{isol},2} = k \ln \frac{N-1}{n-1}$, with 1 subtracted from N in the numerator of the argument of the logarithm because after the first oscillator has been isolated there are 1 fewer total oscillators left in our original System A and in the denominator thereof because there are 1 fewer ground-state oscillators left therein. The negentropy cost of isolating the third of the n ground-state oscillators is $\Delta S_{\text{isol},3} = k \ln \frac{N-2}{n-2}$, of isolating the jth ($1 \leq j \leq n$) $\Delta S_{\text{isol},j} = k \ln \frac{N-(j-1)}{n-(j-1)} = k \ln \frac{N-j+1}{n-j+1}$, of isolating the nth and last $\Delta S_{\text{isol},n} = k \ln \frac{N-n+1}{n-n+1} = k \ln (N-n+1)$. Note that the negentropy cost of isolating ground-state oscillators increases with each one isolated and is highest for the last one isolated. Recalling that T is the temperature of our original System A, the work required to isolate the jth of the n ground-state oscillators is $W_{\text{isol},j} = T\Delta S_{\text{isol},j} = kT \ln \frac{N-j+1}{n-j+1}$. Thus, if at temperature T on average $\langle n \rangle$ of the N harmonic oscillators comprising our original System A are in their ground states, the expectation values of the total negentropy cost $\Delta S_{\text{isol,total}}$ and total work cost $W_{\text{isol,total}} = T\Delta S_{\text{isol,total}}$ of isolating all ground-state oscillators into Subsystem B are, to sufficient accuracy, given by and bounded from above in accordance with:

$$\langle \Delta S_{\text{isol,total}} \rangle = \sum_{j=1}^{\langle n \rangle} \langle \Delta S_{\text{isol},j} \rangle = k \sum_{j=1}^{\langle n \rangle} \ln \frac{N-j+1}{\langle n \rangle - j + 1} < \langle n \rangle \, k \ln (N - \langle n \rangle + 1)$$

$$\implies \langle W_{\text{isol,total}} \rangle = T \langle \Delta S_{\text{isol,total}} \rangle = kT \sum_{j=1}^{\langle n \rangle} \Delta S_{\text{isol},j} = kT \sum_{j=1}^{\langle n \rangle} \ln \frac{N-j+1}{\langle n \rangle - j + 1} < \langle n \rangle \, kT \ln (N - \langle n \rangle + 1) \,. \quad (12)$$

(If $\langle n \rangle$ is not an integer, then the sums in Eq. (12) are, to sufficient accuracy, construed as encompassing all integers j from 1 up through and including the one immediately below $\langle n \rangle$ and then also encompassing the noninteger $\langle n \rangle$.) The inequalities in Eq. (12), bounding $\langle \Delta S_{\text{isol,total}} \rangle$ and $\langle W_{\text{isol,total}} \rangle = T \langle \Delta S_{\text{isol,total}} \rangle$ from above, are justified because the negentropy cost of isolating ground-state oscillators increases with each one isolated and is highest for the last one isolated. Thus even the upper bounds on the negentropy and work costs are modest. The negentropy and work costs computed in Eq. (12) assume thermodynamic perfection (reversibility). But even given typical imperfection (irreversibility), which is inevitable in practice as opposed to in principle, the upper bounds on the actual negentropy and work costs would typically be only a few times larger, and

http://dx.doi.org/10.5772/61641

hence still modest. The Second Law of Thermodynamics requires that the decrease in entropy associated with localizing ground-state oscillators into Subsystem B be paid for by an increase in entropy elsewhere. The payment for any irreversibilities is most typically via waste heat, which must be dumped anywhere except into Subsystem B. It is best dumped into System A's heat reservoir (not into System A itself). The temperature of this reservoir and hence also of System A itself need not be measurably raised if this heat reservoir is very large and/or is comprised of a substance in its two-phase regime. Of course, this waste heat payment will be larger given imperfect (irreversible) than perfect (reversible) operation, but typically only a few times larger.

Note, first, that this is a CSRR (as opposed to TSRR) operation, entailing only *positional* isolation of the oscillators by weight. The isolation and hence *localization* of the n ground-state oscillators into Subsystem B is in the position, not the momentum, part of phase space. Entropy is the logarithmic measure of delocalization and negentropy the logarithmic measure of localization. The negentropy cost for reversible CSRR by weighing is a *localization* cost paid for by work not heat (although the entropy cost exacted owing to any irreversibilities is typically via waste heat, which must be dumped anywhere except into Subsystem B preferably into System A's heat reservoir). Moreover no energy — neither heat nor work — is extracted from either System A or Subsystem B at *any* point during our CSRR process. (Indeed since the oscillators to comprise Subsystem B are in their ground states, energy can*not* be extracted from them.) This is in contrast with both standard and absorption TSRR (recall Sect. 2. and the first paragraph of this Sect.3). Second, since the difference in masses and therefore also weights between an oscillator being in its ground or first excited state is finite, we circumvent the objection that infinitely precise measurements [22] would be required to verify [22] that *precisely* 0 K has been attained. Third, while the unattainability formulation of the Third Law of Thermodynamics forbids the expenditure of the typically small amount of work required to attain 0 K via standard TSRR and the expenditure of the typically small amount of high-temperature heat required to attain 0 K via absorption TSRR, it does *not* forbid the expenditure of the typically small amount of work required to attain 0 K via CSRR. Fourth, we stated "in principle!" — i.e., as a thought experiment — no currently-available or even currently-foreseeable practical weighing technology is sensitive enough. This is in contrast to the QCR systems investigated in Ref. [1], which although more complex, are realizable in practice using currently-available technology.

So does our positional isolation of the n ground-state oscillators into Subsystem B at least *prima facie* seem to challenge the strongest-mode unattainability formulation of the Third Law of Thermodynamics [2–5]? If the unattainability formulation of the Third Law of Thermodynamics in its strongest mode *does* forbid attaining 0 K via CSRR, then it must be for *another reason*. As will be discussed in Sects. 3.3.–3.5, this other reason is *purely* dynamic rather than *thermo*dynamic — the energy-time uncertainty principle.

Of the methods discussed in Ref. [1], the one closest to our weighing thought-experiment discussed in the five immediately preceding paragraphs seems to be that discussed in Sect. 3 of Ref. [1] — but employing only the *first step* of that method. Similarly to the weighing thought-experiment example discussed in the five immediately preceding paragraphs, the proposed real system discussed in Sect. 3 of Ref. [1] (System A by our notation) consists of a mixture of atoms, some of which are in the ground state and some in the first excited state. Also as in our weighing thought-experiment, the temperature is assumed low enough so that the probability of occupancy of the second or higher excited states is negligible. This first step of the method employed in Sect. 3 of Ref. [1] entails positional isolation of atoms on the

ground state from those in the first excited state by a Stern-Gerlach apparatus. The subsystem comprised of atoms in the ground state after positional isolation via the Stern-Gerlach apparatus constitutes Subsystem B, our subsystem at the absolute zero of temperature, 0 K. This positional isolation of atoms is a CSRR process. This Stern-Gerlach-apparatus version of our thought experiment may, in accordance with Sect. 3 of Ref. [1], be more realizable experimentally than the weighing version thereof.

In contrast with Sect. 3 of Ref. [1], in the Stern-Gerlach modification of our weighing CSRR method no attempt is made to thence also de-excite the atoms in the first excited state down to the ground state, the *second step* of the QCR method discussed in Sect. 3 of Ref. [1]. It is important to recognize that this *second step* is *neither* a TSRR process *nor* a CSRR process. The entropy of a set of atoms is zero if they are *all in the same quantum state*, irrespective of whether this quantum state is the ground state or not. (But see the last paragraph of the Appendix concerning this point.) In the particular case currently under consideration, we have a set of atoms *all in the first excited state*. Their de-excitation from the first excited state to the ground state maintains their entropy constant at zero. Thus it is *not* an entropy-reduction (SR) process — it is neither a TSRR process nor a CSRR process. It is rather *another type* of refrigeration process, which we have dubbed as *energy*-reduction refrigeration (ERR): *energy* E but *not* entropy S is extracted to de-excite these atoms from the first excited state to the ground state. Thus, as was the case with one-time-expansion ERR which we considered in Sect. 2, especially in Sect. 2.5, this version of ERR yields rather than costs work. But unlike one-time-expansion ERR which we considered in Sect. 2.1, especially in Sect. 2.5, this version of ERR process does *not* require infinite *volume* to attain 0 K. Moreover, the extracted energy E is *non*thermal, because the oscillators are initially all in the first excited state, *not* in a Boltzmann distribution among states. Similarly as is the case with respect to the expenditure of the typically small amount of work required to attain 0 K via CSRR, the unattainability formulation of the Third Law of Thermodynamics does *not* forbid ERR by forbidding the extraction of the typically small amount of *non*thermal energy required to de-excite these atoms from the first excited state to the ground state. Thus if the unattainability formulation of the Third Law of Thermodynamics in its strongest mode *does* forbid ERR, then, as with CSRR, it must be for *another reason*. As will be discussed in Sects. 3.3.–3.5, this other reason is, as with CSRR, *purely* dynamic rather than *thermo*dynamic — the energy-time uncertainty principle, which imposes the requirement of infinite *time* to attain 0 K.

But, for the moment, not considering the energy-time uncertainty principle, once it has been established and proven that via a CSRR process, entailing isolation in position space rather than in momentum space, e.g., via weighing or employment of a Stern-Gerlach apparatus, that we can be *perfectly* sure that *all* n oscillators in our new ground-state-only Subsystem B really are in the ground state, then this system is indeed at *precisely* 0 K. By contrast, *even* not considering the energy-time uncertainty principle, we can never be *perfectly* sure that all N oscillators in our original System A really are in the ground state so long as System A's temperature $T > 0$ is positive, however slightly positive [2–5]. No matter how slightly positive, we can never be *perfectly* sure that all N System-A oscillators are in their ground states. In explanation, let the N oscillators in our original System A be in thermal equilibrium with a heat reservoir at positive temperature T so small that the probability of any one given System-A oscillator being in its first excited state is $\mathbb{P}_{A1} \doteq e^{-\Delta E/kT} \lll 1$, and we can neglect the probability of even one of them being in its second excited or higher excited state. Thus, the probability that not even one System-A oscillator is in the first excited state, and hence that all N of them are in the ground state, is $\mathbb{P}_{A0}^{N} = (1 - \mathbb{P}_{A1})^{N} \doteq \left(1 - e^{-\Delta E/kT}\right)^{N}$, which

http://dx.doi.org/10.5772/61641

for arbitrarily small positive T and N *not* arbitrarily large simplifies to $\mathbb{P}_{A0}^{N} \doteq 1 - Ne^{-\Delta E/kT}$. For arbitrarily small positive T, if N is *not* arbitrarily large, $\mathbb{P}_{A0}^{N}$ can be arbitrarily close to 1, but it can never be *precisely* 1 as is required to attain *precisely* 0 K. If N *is* arbitrarily large, then the situation is even worse. For then, however large ΔE and however small T, it is certain that at least one System-A oscillator is in its first excited state [28]. But this is a limitation *only* of our *original* System A of N oscillators, *not* of our Subsystem B of n oscillators in their ground state that we have positionally isolated by weighing, by a Stern-Gerlach apparatus, or by any other CSRR method. For arbitrarily small positive T, if N is *not* arbitrarily large, $\langle n \rangle$ can be considerably closer to N than to $N - 1$, so that if we form Subsystem B it would likely — but *not* for sure — contain all N oscillators of System A. The "*not*" in "*not* for sure" is why, if T is positive, however slightly positive, System A is *not* our new ground-state-only Subsystem B.

Positional isolation via pure CSRR is not the only method by which absolute zero might be attained. The attainment of absolute zero via QCR methods [1], of which the specific one discussed in Sect. 3 of Ref. [1] employs first CSRR and then ERR, has been investigated [1]. But positional isolation via pure CSRR seems simpler and easier in principle, even if, via weighing, it may not be realizable in practice. But perhaps as discussed three paragraphs previously, via a Stern-Gerlach apparatus it may be [1]. Its simplicity in principle allows us to focus on attainment of absolute zero per se rather on experimental technical issues. Moreover, as noted three paragraphs previously, the QCR method discussed in Sect. 3 of Ref. [1] employs purely-CSRR positional isolation as its first step; as noted two paragraphs previously, only the de-excitation of atoms still in the first excited state down to the ground state in its second step is an ERR process.

3.3. The energy-time uncertainty principle: a *purely* dynamic (*not thermo*dynamic) Third-Law limitation under quantum mechanics

We re-emphasize (recall Sect. 3.2, especially the second-to-last paragraph thereof) that the attainment of absolute zero requires *perfect* certainty that our *entire* new n-oscillator Subsystem B is in its ground state — that *all* n oscillators of Subsystem B are in the ground state. But the energy-time uncertainty principle may contravene [29–41]. [Dr. Bernard L. Cohen [29] employs the energy-time uncertainty principle in discussing quantum fluctuations. Dr. Robert Gomer [30] (cited by Dr. Cohen [29]) shows how the position-momentum uncertainty principle can be employed in more limited circumstances. Dr. Mark J. Hagmann [31–35] extends and evaluates Dr. Cohen's work, and compares it with other works. Drs. Donald H. Kobe and V. C. Aguilera-Navarro [38] provide a derivation from first principles of the energy-time uncertainty relation [38], which they and Drs. Hiromi Iwamoto and Mario Goto employ in a study of tunneling times [39]. Drs. V. V. Dodonov and A. V. Dodonov provide extensive considerations concerning the energy-time uncertainty principle [40]. A heuristic overview is provided by the current author [41].] In order to ensure that *all* n oscillators in Subsystem B really are in the ground state, the best that the energy-time uncertainty principle allows us to do is to isolate each of these oscillators for a sufficiently long time interval Δt pursuant to its being incorporated into Subsystem B, or equivalently to isolate Subsystem B for a sufficiently long time Δt. Let us estimate how long Δt must be. Recall that we let ΔE denote an energy *gap* — in the case currently under consideration the energy *gap* between adjacent states of any given one of our harmonic oscillators. Let $\Delta\mathcal{E}$ denote the magnitude of a *quantum* energy *fluctuation*, and let $\Delta\mathfrak{E}$ denote

the magnitude of a *thermal* energy *fluctuation*. The minimum possible root-mean-square *quantum* fluctuation magnitude that the energy-time uncertainty principle allows in energy $\Delta\mathcal{E}_{\rm rms}$ during a time interval Δt is $\Delta\mathcal{E}_{\rm rms} = \hbar/2\Delta t$ [29–41]. (Spontaneously-occurring quantum fluctuations are, or at least tend to be, of minimal possible magnitude, as is required for the macroscopic world being maximally close to classical [29–41].) We require $\Delta\mathcal{E}_{\rm rms}$ to be much smaller than the typical *upper* limiting root-mean-square magnitude $\Delta\mathfrak{E}_{\rm rms} \approx kT$ of *thermal* energy fluctuations in our original System A. $\Delta\mathfrak{E}_{\rm rms} \approx kT$ is a typical *upper* limiting thermal-energy-fluctuation root-mean-square magnitude because if $N - \langle n\rangle \ll N$ then most oscillators in System A will usually be in the ground state and hence at least $\Delta E \gg kT$ and hence at most $\Delta\mathfrak{E}_{\rm rms} \lesssim kT$ and more to be expected $\Delta\mathfrak{E}_{\rm rms} \ll kT$. But we also require $\Delta E \gg kT$ for the energy gap between harmonic-oscillator energy levels to ensure that probability that the second or higher excited states are occupied can be neglected compared to the already small probability that the first excited state is occupied (recall the first paragraph of Sect. 3.2). Thus all told we require $\Delta\mathcal{E}_{\rm rms} = \hbar/2\Delta t \ll \Delta\mathfrak{E}_{\rm rms} \lesssim kT \ll \Delta E$. This implies that the strong inequality $\Delta t \gg \hbar/2\Delta\mathfrak{E}_{\rm rms} \gtrsim \hbar/2kT$ and the even stronger one $\Delta t \ggg \hbar/2\Delta E$ must be fulfilled [29–41]. But even such a long Δt is not long enough to allow us to be *perfectly* certain that all n oscillators in our new ground-state-only Subsystem B really are in the ground state, but only to be *almost* perfectly certain that they are [29–41]. Thus, our caveat is as follows: In order to be *perfectly* certain that all n oscillators in our new ground-state-only Subsystem B really are in the ground state, each oscillator must be isolated for $\Delta t \to \infty$ — for infinite time, forever — pursuant to its being incorporated into our new ground-state-only oscillator Subsystem B, or equivalently Subsystem B must be isolated for $\Delta t \to \infty$. The quantum-mechanical probability $\mathbb{P}_{B1}$ that any one given oscillator isolated for inclusion in our new Subsystem B is in its first excited state decays exponentially or at least quasi-exponentially with increasing Δt in accordance with [29–41]

$$\begin{aligned} &\mathbb{P}\left(\Delta\mathcal{E}\Delta t\right) \sim e^{-\Delta\mathcal{E}/\Delta\mathcal{E}_{\rm rms}} = e^{-2\Delta\mathcal{E}\Delta t/\hbar} \\ &\overset{\Delta\mathcal{E}=\Delta E}{\Longrightarrow} \mathbb{P}_{B1} = \mathbb{P}\left(\Delta E\Delta t\right) \sim e^{-\Delta E/\Delta\mathcal{E}_{\rm rms}} = e^{-2\Delta E\Delta t/\hbar}. \end{aligned} \tag{13}$$

(This exponential or at least quasi-exponential decay is brought out in Ref. [29] and is important implicitly and/or explicitly in Refs. [30–35] and [41]. It is not specifically mentioned but is not inconsistent with Refs. [36–40].) The first line of Eq. (13) expresses the general approximate probability of a quantum energy fluctuation of magnitude $\Delta\mathcal{E}$ persisting for time Δt. In the second line of Eq. (13) $\Delta\mathcal{E}$ is set equal to the energy gap ΔE between adjacent harmonic-oscillator energy states (the gap between the ground and first excited states being of current interest). Thus the probability that any one given Subsystem-B oscillator is in its ground state after isolation for Δt is $\mathbb{P}_{B0} = 1 - \mathbb{P}_{B1} = 1 - \mathbb{P}\left(\Delta E\Delta t\right) \sim 1 - e^{-2\Delta E\Delta t/\hbar}$. Hence the probability that all n Subsystem-B oscillators are in their ground states after isolation for Δt is $\mathbb{P}_{B0}^n = \left(1 - \mathbb{P}_{B1}\right)^n = \left[1 - \mathbb{P}\left(\Delta E\Delta t\right)\right]^n \sim \left(1 - e^{-2\Delta E\Delta t/\hbar}\right)^n$, which for n not too large and sufficiently large Δt simplifies to $\mathbb{P}_{B0}^n \sim 1 - ne^{-2\Delta E\Delta t/\hbar}$. Thus $1 - \mathbb{P}_{B0}^n \sim ne^{-2\Delta E\Delta t/\hbar}$, which depends only *linearly* on n but decreases *exponentially* (or at least quasi-exponentially) with increasing Δt, soon becomes negligible. But however strongly negligible it becomes, it *never* becomes *precisely* 0 except in the limit $\Delta t \to \infty$. And $1 - \mathbb{P}_{B0}^n$ is required to be *precisely* 0 — equivalently $\mathbb{P}_{B0}^n$ is required to be *precisely* 1 — if *precisely* 0 K is to be attained and if we are to have *perfect* verification [22] that *precisely* 0 K has been attained.

http://dx.doi.org/10.5772/61641

The difference in masses and therefore also weights of an oscillator being in its ground state as opposed to in its first excited state is finite, thereby, as per the first five paragraphs of Sect. 3.2, circumventing the objection that infinitely precise measurements [22] would be required to verify [22] that *precisely* 0 K has been attained. This objection is similarly circumvented if instead of weighing we employ a Stern-Gerlach apparatus as per the sixth paragraph of Sect. 3.2, in accordance with the first step of the method employed in Sect. 3 of Ref. [1]. But the objection posed by the energy-time uncertainty principle seems to be uncircumventable: *Exact* attainment of 0 K and *perfect* verification [22] that *precisely* 0 K has been attained seems to require infinite time. Thus, the energy-time uncertainty principle may provide additional — quantum-mechanical and hence *purely* dynamic as opposed to *thermo*dynamic — protection against *exact* attainment of 0 K and *perfect* verifiability [22] that *precisely* 0 K has been attained, and hence against the unattainability formulation of the Third Law of Thermodynamics in its strongest mode being *precisely* violated. It is not clear whether or not the energy-time uncertainty principle imposes a similar limitation on the QCR systems and methods discussed in Ref. [1]. But owing to the universality of quantum mechanics and hence of the energy-time uncertainty principle, this seems likely to be the case. Indeed owing to the universality of quantum mechanics and hence of the energy-time uncertainty principle, this seems likely to be the case in general, irrespective of the refrigeration method — TSRR, CSRR, ERR, QCR, etc., or any combination thereof — that is employed. This is in accordance with the conclusion reached in Refs. [6–8] via far more technical and mathematical analyses.

Note the *qualitative* — not merely *quantitative* — distinction between the *thermo*dynamic (Boltzmann-distribution) probability $\mathbb{P}_A$ discussed in Sect. 3.2. as opposed to the *purely* dynamic (quantum-mechanical) probability $\mathbb{P}_B$ discussed in this Sect. 3.3. *Even if*, *thermo*dynamically, *exact* attainment of 0 K and *perfect* verification [22] that *precisely* 0 K has been attained *could be* achieved for Subsystem B, the *pure* dynamics of quantum mechanics, specifically the energy-time uncertainty principle, seems to impose the requirement that infinite time must elapse first. [This distinction between *thermo*dynamic probabilities as opposed to *purely* dynamic (quantum-mechanical) probabilities should not be confused with the distinction between the derivation of the *thermo*dynamic Boltzmann distribution per se in classical as opposed to quantum statistical mechanics. The latter distinction, which we do not consider in this chapter, obtains largely owing to the postulate of random phases being required in quantum but not classical statistical mechanics [42,43].]

Nevertheless, given the *exponential* (or at least quasi-exponential) decay of $1 - \mathbb{P}^n_{B0} \sim ne^{-2\Delta E \Delta t/\hbar}$ [29–41], fulfillment of the *very* strong inequality $\Delta t \ggg \hbar/2\Delta E$ *does* seem to imply that we can be close enough to perfectly certain that all n oscillators in our new Subsystem B really are in the ground state *for all practical purposes*. Hence it seems that we must be content with attainment for all practical purposes as opposed to exact attainment of 0 K and verification for all practical purposes as opposed to perfect verification [22] that *precisely* 0 K has been attained. Thus perhaps the energy-time uncertainty principle provides the ultimate protection against *perfect* violation of the unattainability formulation of the Third Law of Thermodynamics in its strongest mode. But given the *exponential* (or at least quasi-exponential) decay of $1 - \mathbb{P}^n_{B0} \sim ne^{-2\Delta E \Delta t/\hbar}$ [29–41], perhaps, while not *perfectly* violating the strongest-mode unattainability formulation of the Third Law of Thermodynamics, CSRR as opposed to TSRR at least challenges it in the strongest manner that the laws of physics allow. Recall from the sixth and seventh paragraphs of Sect. 3.2. that the specific QCR method discussed in Sect. 3 of Ref. [1] employs first CSRR and then ERR [1], and that we employed the first step of this method for CSRR positional isolation

via Stern-Gerlach apparatus. Recall also that in the last two paragraphs of Sect. 2.3. we have already employed exponential decay in consideration of rendering insulation for a refrigerated system perfect *for all practical purposes*, even if it cannot be *exactly* perfect [23].

3.4. The quest for absolute zero under classical versus quantum mechanics

The energy-time uncertainty principle is of purely quantum-mechanical origin. It does not exist in classical mechanics, whether Newtonian or relativistic. Thus under classical mechanics, whether Newtonian or relativistic, it might seem, at least *prima facie*, that, at least in principle, CSRR via positional isolation by means of weighing as discussed in the first five paragraphs of Sect. 3.2, by means of a Stern-Gerlach apparatus as discussed in the sixth paragraph of Sect. 3.2 (whether or not enhanced via ERR as per the seventh paragraph of Sect. 3.2), or via QCR in general [1], *can* not only attain *precisely* 0 K but also provide *perfect* verification [22] that *precisely* 0 K has been attained — and both in *finite*, even *arbitrarily short*, time Δt. Hence it may seem, at least *prima facie*, that under classical mechanics, at least in principle, the unattainability statement of the Third Law of Thermodynamics even in its strongest mode *can* be *precisely* violated via CSRR. However, experimentally realizable proposals for attaining 0 K are quantum-mechanical [1]. Indeed the entire Universe is ultimately quantum-mechanical; classical mechanics, Newtonian or even relativistic, being only a limiting approximation. Hence attainment of 0 K *for all practical purposes* and verification [22] that 0 K has been attained *for all practical purposes* is probably the best that can be achieved. Our conclusion seems unalterable even if one accepts the viewpoint expressed by Dr. David Bohm that classical mechanics should be considered in its own right and as prerequisite for quantum mechanics, rather than as a limiting case of quantum mechanics [44]. This is opposed to the more generally accepted viewpoint that classical mechanics should be considered as a limiting case of quantum mechanics. Moreover, even Dr. Bohm expresses the latter viewpoint in his recognition of the Universe as being ultimately quantum-mechanical [45].

Besides, classical mechanics imposes its own burden on the quest for absolute zero. While there are instances of quantization even in classical mechanics, for example the discrete allowed frequencies and wavelengths of a vibrating string of finite length [46], of electromagnetic waves within a finite volume [47–49], and of sound waves within a finite volume [47–49], so far as is known energy is always continuous and never quantized in classical mechanics.[4] Quantization of frequency ν and hence also of wavelength λ, for example the discrete allowed frequencies and wavelengths of a vibrating string of finite length [46], of electromagnetic waves within a finite volume [47–49], and of sound waves within a finite volume [47–49], implies quantization of energy E in *quantum* mechanics in accordance with the *quantum*-mechanical relation $E = h\nu = hc/\lambda$ [50–52]. (Here c is the speed of wave prorogation, whether of waves on a string, of light waves, of sound waves, etc. Note for example that for photons ν in a material medium equals ν in a vacuum; both c and λ are smaller in a material medium than in a vacuum by a ratio equal to the index of refraction of the medium.) The relation $E = mc^2 = h\nu = hc/\lambda$ is at the heart of the very closely related Einstein [50–52] and deBroglie [50–52] postulates [50–52]. The relation $E = mc^2$ obtains

[4] (Re: Entries [48] and [49], Refs. [48] and [49]) Photons of equilibrium blackbody radiation are discussed in Sect. 10.6 and phonons of sound waves in solids in Problem 10.8 of Chap. 10 on pp. 369–371 of Ref. [48]. Both photons and phonons are discussed in Chap. 6 of Ref. [49].

http://dx.doi.org/10.5772/61641

in both classical and quantum relativistic mechanics. But quantization of frequency ν and hence also of wavelength λ does *not* imply quantization of energy E in *classical* mechanics because the relation $E = h\nu = hc/\lambda$ is *strictly quantum*-mechanical [50–52]. There exists *no classical*-mechanical relation such as $E = h\nu = hc/\lambda$ [50–52]. Even in quantum mechanics the relation $E = h\nu = hc/\lambda$ is necessary but not sufficient for discreteness of energy levels as opposed to an energy continuum, for in an infinite volume ν and λ can take on a continuum of values. But given only the additional very mild condition of a finite volume — a fixed finite volume or even an unfixed but always finite volume for example corresponding to maintenance of constant pressure — only discrete values of λ and hence of $\nu = c/\lambda$ will fit therein, thus ensuring discreteness of energy levels under quantum — but not classical — mechanics. This can also be shown via considerations of the Schrödinger equation [53]. Thus so far as is known quantization of energy [50–52] and discrete energy levels [53] can exist under quantum mechanics [50–53], indeed *must* exist under quantum mechanics given finite volume, but can*not* exist under classical mechanics [50–53]. Hence under classical mechanics, owing to continuity of energy, infinitely precise measurements [2,3,22] would be required to *perfectly* verify [2,3,22] that *precisely* 0 K has been attained [2,3,22]: With an infinitesimal gap between the ground and first excited states, weighing of our harmonic oscillators would have to be infinitely precise. With an infinitesimal gap between the ground and first excited states of atoms an infinitely-sensitive Stern-Gerlach apparatus would be required to separate atoms in the two states. By contrast, under quantum mechanics, owing to discreteness of energy levels [50–53] of a system within a finite volume, measurements of merely finite precision suffice for verification [22] that *precisely* 0 K has been attained.

Thus quantum mechanics, via the energy-time uncertainty principle, imposes the requirement of isolation for infinite time for *perfect* verification [22] that *precisely* 0 K has been attained, but by requiring discreteness of energy levels for a system of finite volume lifts the requirement of infinitely precise measurements [22] to *perfectly* verify [22] that *precisely* 0 K has been attained. By contrast, classical mechanics, since it lacks an energy-time uncertainty principle, lifts the requirement of isolation for infinite time for *perfect* verification [22] that *precisely* 0 K has been attained, but by requiring an energy continuum imposes the requirement of infinitely precise measurements [22] to *perfectly* verify [22] that *precisely* 0 K has been attained. Of these two requirements, the first seems less onerous than the second, because as discussed in Sect. 3.3. the *un*certainty that *precisely* 0 K has been attained decays *exponentially* (or at least quasi-exponentially) with time [29–41]. This exponential or at least quasi-exponential decay mitigates (albeit does not remove) the requirement of isolation for infinite time under quantum mechanics for *perfect* verification [22] that *precisely* 0 K has been attained. No such decay, exponential, quasi-exponential, or otherwise, mitigates the requirement under classical mechanics for an energy continuum, hence implying that infinitely precise measurements [22] are requisite for perfect verification [22] that *precisely* 0 K has been attained. So at least *prima facie* it seems that quantum mechanics at least brings us closer than classical mechanics to *perfect* verifiability [22] that *precisely* 0 K can be attained.

3.5. Summary: TSRR versus CSRR

In summary, the *thermo*dynamic difficulties in *attaining precisely* 0 K via TSRR [2–5] seem to be circumventable via CSRR. By contrast, the *purely* dynamic (quantum-mechanical) limitation imposed by the energy-time uncertainty principle as per Sects. 3.3. and 3.4. is, strictly, not circumventable via either TSRR or CSRR, but this limitation may not be crucial if we do

not insist on *exact* attainment of 0 K and *perfect* verification [22] that *precisely* 0 K has been attained, but are content with attainment of 0 K and verification [22] of its attainment that is perfect *for all practical purposes*.

Thus it seems that CSRR, based on localization of refrigerated entities in position space, as discussed in Sect. 3, at least brings us *closer* to *exact* attainment of $T_C = 0\,\mathrm{K}$ and to *perfect* verification [22] that 0 K has been attained than does TSRR, based on localization of refrigerated entities in momentum space, as discussed in Sect. 2. And CSRR under quantum mechanics (e.g., via weighing or via Stern-Gerlach apparatus), or via QCR methods [1], as opposed to under classical (Newtonian or relativistic) mechanics, seems to bring us the *closest*. (Recall from the sixth and seventh paragraphs of Sect. 3.2. that the specific QCR method discussed in Sect. 3 of Ref. [1] employs first CSRR via Stern-Gerlach apparatus and then ERR [1]; we employed the first, CSRR, step thereof in the sixth paragraph of Sect. 3.2.)

Moreover, *even if* $T_C = 0\,\mathrm{K}$ *could* be precisely *attained* and also its attainment *could* be *perfectly* verified [22], whether via standard or absorption TSRR, via CSRR, via one-time-expansion ERR as discussed in Sect. 2, via ERR as part of a QCR method as discussed in this Sect. 3, or via other ERR methods, via QCR, etc., or via any combination thereof, the question of *maintaining* $T_C = 0\,\mathrm{K}$ as per Sect. 2.3. and the last paragraphs of Sects. 2.4. and 2.5. is still open [23]. *Even if* $T_C = 0\,\mathrm{K}$ *could* be precisely *attained* and also its attainment could be *perfectly* verified [22], the question of whether or not it can be *maintained* for finite time, as opposed to merely the infinitesimally short time allowed in accordance with Eqs. (6) and (7) and the associated discussions, is still open. But least in principle even if not in practice *if* $T_C = 0\,\mathrm{K}$ can be *attained*, whether *maintainable* only for an instant or longer, then this may be more easily achievable via CSRR, especially under quantum rather than classical mechanics, than via standard or even absorption TSRR (or any other TSRR). The same is probably true with respect to QCR as opposed to TSRR.

But the issue of *maintenance* [23] seems *inseparable* from that of *verifiability* [22]. For, as discussed in Sects. 3.3. and 3.4, the energy-time uncertainty principle requires $\Delta t \to \infty$ for perfect verifiability that $T_C = 0\,\mathrm{K}$ has been attained, which is obviously incompatible with $T_C = 0\,\mathrm{K}$ being maintained only for an instant, or even for any finite number of instants, in accordance with Eqs. (6) and (7) and the associated discussions. If $T_C = 0\,\mathrm{K}$ can be maintained only for an instant, (or any finite number of instants), then the energy-time uncertainty principle seems to preclude verification *even for all practical purposes* that $T_C = 0\,\mathrm{K}$ has actually been attained. Thus verification [22] *even for all practical purposes* that $T_C = 0\,\mathrm{K}$ has actually been attained seems to require insulation [23] that is perfect *for all practical purposes* as per Sect. 2.3.

3.6. What if: Better-than-perfect refrigeration?

The Second Law of Thermodynamics forbids better-than-perfect refrigeration, in which the total entropy change is negative. Yet the universal validity of the Second Law of Thermodynamics has been seriously questioned [54–58], albeit with the understanding that even if not universally valid at the very least it has a very wide range of validity [54–58]. Thus *what if* better-than-perfect refrigeration *is* possible, whether via TSRR, CSRR, ERR, QCR, etc., or any combination thereof? (ERR entails zero entropy change; hence it could be part, but not the entirety, of a better-than-perfect refrigeration process, if such can exist [54–58]. Recall from the sixth and seventh paragraphs of Sect. 3.2. that the specific

http://dx.doi.org/10.5772/61641

QCR method discussed in Sect. 3 of Ref. [1] employs first CSRR and then ERR [1]: we employed the first, CSRR, step thereof in the sixth paragraph of Sect. 3.2.) For example, what if CSRR-isolation of our Subsystem-B oscillators could be achieved at smaller negentropy and work cost than in accordance with Eq. (12)? Unfortunately, *even if* better-than-perfect refrigeration in contravention of the Second Law of *Thermo*dynamics is possible, the *purely* dynamic limitations discussed in Sects. 3.3.–3.5. still obtain for both TSRR and CSRR (as well as for ERR, QCR, etc., or any combination of refrigeration methods). Thus at least *prima facie* it seems that *even if* the second law can be contravened [54–58] and hence better-than-perfect refrigeration is possible, owing to the energy–time uncertainty principle the strongest-mode unattainability statement of the third law could still be violated only for all practical purposes and not perfectly. And this considers *only* the difficulties of *attaining* $T_C = 0\,\mathrm{K}$. *Even if* the Second Law of Thermodynamics can be contravened [54–58] and hence better-than-perfect refrigeration is possible, the difficulties in *maintaining* $T_C = 0\,\mathrm{K}$ for longer than an infinitessimal time, discussed in Sect. IIC, the last paragraphs of Sects. 2.4. and 2.5, and the last two paragraphs of Sect. 3.5, may preclude verification [22] *even for all practical purposes* that $T_C = 0\,\mathrm{K}$ has actually been attained — unless insulation [23] that is perfect *for all practical purposes* as per Sect. 2.3. is possible.

4. Hot true and cold effective negative Kelvin temperatures

4.1. Hot true negative Kelvin temperatures

Negative Kelvin temperatures certainly exist [59–63]. But *true* negative Kelvin temperatures are *hotter* than $T = \infty\,\mathrm{K}$, not colder than $T = 0\,\mathrm{K}$ [59–62]. *True* negative Kelvin temperatures exist only in systems with an upper bound in energy, wherein $T = (\partial E/\partial S)_{V,N} < 0$ obtains if enough energy is pumped into such a system so that its high-energy state is more populated than its low one — a population inversion [59–62]. [As per standard notation, the subscript "V, N" denotes fixed volume and number of entities (most typically atoms or molecules).] The temperature of *any* system can, at least in principle, be raised to $T = \infty\,\mathrm{K}$ via energy pumped into the system as heat and/or as work. But unless a heat reservoir at a negative Kelvin temperature is available, energy must be pumped into a system as work rather than as heat, i.e., *non*thermally, if the system's temperature is to be raised to negative Kelvin values, because heat input from a heat reservoir at a positive Kelvin temperature can never raise a system's temperature above $T = \infty\,\mathrm{K}$, which corresponds to all of the system's states being uniformly populated — just short of a population inversion [59–62].

Consider, for simplicity, a 2-energy-level system with both levels nondegenerate. A total of N entities (typically atoms whose nuclei can manifest spin aligned either parallel or antiparallel to an external magnetic field) can be distributed among these 2 energy levels. At $T = +0\,\mathrm{K}$, the probability is unity that all N entities are in the lower level and hence the system's entropy is minimized at $S = 0$. As energy is pumped into the system (as heat and/or as work), its temperature $T = (\partial E/\partial S)_{V,N}$ increases through increasing positive values and its entropy increases. At $T = +\infty\,\mathrm{K} = -\infty\,\mathrm{K}$, each entity has a probability of $1/2$ of being in either level and hence the system's entropy is maximized at $S = Nk\ln 2$. As more energy is pumped into the system (as work only unless a heat reservoir at $T = -0\,\mathrm{K}$ is available), its temperature $T = (\partial E/\partial S)_{V,N}$ increases through decreasing negative values from $T = -\infty\,\mathrm{K}$ to $T = -0\,\mathrm{K}$ and its entropy decreases, until at $T = -0\,\mathrm{K}$ the probability is unity that all N entities are in the upper level and hence the system's entropy is again minimized at $S = 0$.

It should be noted that the concept of hot negative Kelvin temperature can meaningfully be applied for 2-energy-level systems [63], whether or not either level or both are degenerate — but *only* for 2-energy-level systems [63]. For systems with 3 or more energy levels, wherein population need not be a monotonic function of level (multiple population inversions are possible with 4 or more levels), the concept of hot negative Kelvin temperature becomes unwieldy and contrived [63].[5]

It has been remarked [59–62] that there are advantages in defining temperature via $1/T = -(\partial S/\partial E)_{V,N}$, because by this definition the numerical value of a system's temperature always increases monotonically with its increasing ability to spontaneously deliver heat to its surroundings or equivalently with its decreasing ability to spontaneously accept heat from its surroundings, whether temperature defined via $T = (\partial E/\partial S)_{V,N}$ is positive or negative [59–62]. But temperature defined via $T = (\partial E/\partial S)_{V,N}$ has the advantages of numerical proportionality to temperature as measured by an ideal-gas thermometer and to average thermal kinetic energy per degree of freedom of ideal-gas molecules in the classical (nonquantum) regime. So we employ the definition $T = (\partial E/\partial S)_{V,N}$ in this chapter.

4.2. Cold effective negative Kelvin temperatures

Insofar as is known, *true* negative Kelvin temperatures that are *colder* than $T = 0\,\mathrm{K}$ do not exist. Nevertheless, we can still consider *effective* negative Kelvin temperatures that are *colder* than $T = 0\,\mathrm{K}$ — linearly extrapolating the Kelvin temperature scale downwards through $T = 0\,\mathrm{K}$ to negative *effective* values. Such cold *effective* negative Kelvin temperatures *do* exist. Consider, for example, the effective wind-chill temperature $\mathcal{W}$ on Neptune, at the level in Neptune's atmosphere where the pressure is 1 bar, approximately 1 atm. The wind-chill temperature $\mathcal{W}$ is the temperature that calm air must have to produce the same chilling effect as moving air — wind — at speed $\mathcal{V}$, all other things being equal. The *true* mean temperature (without wind chill) at the 1 bar level on Neptune is approximately $T = 72\,\mathrm{K} = -201\,^\circ\mathrm{C} = -330\,^\circ\mathrm{F}$ [65]. The standard *formula* for wind-chill temperature $\mathcal{W}$ employed by the U. S. A. National Weather Service is [66]

$$\mathcal{W} = \left[0.6215 T_{\circ\mathrm{F}} + (0.4275 T_{\circ\mathrm{F}} - 35.75)\, \mathcal{V}_{\mathrm{mi/h}}^{0.16} + 35.74\right]^\circ\mathrm{F}\,. \tag{14}$$

In Eq. (14), the wind speed $\mathcal{V}$ is that at the 5 ft (typical face) level, based on reduction owing to surface friction of wind speed measured at the standard 10 m or 33 ft level to the 5 ft level [66]. But over flat open ground or over open water, this reduction in wind speed is

[5] (Re: Entry [63], Ref. [63]) In Ref. [63], Dr. Peter Atkins doesn't seem to explicitly state that negative Kelvin temperatures are hotter than $\infty\,\mathrm{K}$, not colder than 0 K. He admits the possibility of attaining 0 K via noncyclic processes, but as we showed in Sect. 3. of this chapter *purely* dynamic — as opposed to *thermo*dynamic — limitations may contravene. On pp. 103–104 of Ref. [63], he correctly states that the third law of thermodynamics is "not really in the same league" as the zeroth, first, and second laws, and that "hints of the Third Law of Thermodynamics are already present in the consequences of the second law," but that the Third Law of Thermodynamics is "the final link in the confirmation that Boltzmann's and Clausius's definitions refer to the same property." But his statement that "we need to do an ever increasing, and ultimately infinite, amount of work to remove energy from a body as heat as its temperature approaches absolute zero" neglects the rapid decrease in specific heat as absolute zero is approached as discussed in Sect. 2. of this chapter.

http://dx.doi.org/10.5772/61641

typically small. So for simplicity let us neglect this typically small difference in wind speeds.[6] For the given temperature $T = 72\,\mathrm{K} = -201\,^\circ\mathrm{C} = -330\,^\circ\mathrm{F}$ at the 1 bar level on Neptune [65], even with a slow (by Neptune standards) $\mathcal{V} = 50\,\mathrm{mi}\,/\,\mathrm{h}$ wind, Eq. (14) yields $\mathcal{W} = -500\,^\circ\mathrm{F} = -296\,^\circ\mathrm{C} = -22\,\mathrm{K}$. [A wind speed of $\mathcal{V} = 50\,\mathrm{mi}\,/\,\mathrm{h}$ is chosen so that our example is more "Earthlike." Typical wind speeds on Neptune are considerably higher than 50 mi / h (See Ref. [65].) But according to Eq. (14), the chilling effect of wind increases at a decreasing rate with increasing wind speed: $(\partial\mathcal{W}/\partial\mathcal{V})_T = \left[0.16\,(0.4375T_{\circ\mathrm{F}} - 35.75)\,\mathcal{V}_{\mathrm{mi}\,/\,\mathrm{h}}^{-0.84}\right]\,^\circ\mathrm{F}\,/\,(\mathrm{mi}\,/\,\mathrm{h})$. Hence $\mathcal{W}$ decreases only very slowly at wind speeds above 50 mi / h, at least assuming that if not Eq. (14) in its entirety then at least this aspect of Eq. (14) retains at least approximate validity at Neptune-like temperatures. The singularity in $(\partial\mathcal{W}/\partial V)_T$ at $\mathcal{V} = 0\,\mathrm{mi}\,/\,\mathrm{h}$ is sufficiently weak that it has no effect on values of $\mathcal{W}$ itself.] Since standard atmospheric pressure at sea level on Earth is approximately 1 bar, for illustrative purposes and for argument's sake let us assume that the standard wind chill *formula* [Eq. (14)] retains at least approximate validity at the 1 bar level on Neptune, especially since the atmospheric density of $0.45\,\mathrm{kg}\,/\,\mathrm{m}^3$ at the 1 bar level on Neptune is at least comparable to that at the 1 bar level on Earth. (We will appraise this assumption later in this Sect. 4.2, especially in the second-to-last paragraph thereof.) The temperature in Neptune's atmosphere at the 0.1 bar level is $T = 55\,\mathrm{K} = -218\,^\circ\mathrm{C} = -361\,^\circ\mathrm{F}$ [65]. Since Eq. (14) was derived for standard conditions (1 bar atmospheric pressure on Earth), its accuracy may be reduced if it is applied at the 0.1 bar level on Neptune. If we nevertheless apply it at the 0.1 bar level on Neptune, we obtain, even with a slow (by Neptune standards) $\mathcal{V} = 50\,\mathrm{mi}\,/\,\mathrm{h}$ wind, $\mathcal{W} = -544\,^\circ\mathrm{F} = -320\,^\circ\mathrm{C} = -47\,\mathrm{K}$.

The standard wind-chill *formula* [Eq. (14)] should not be confused with the standard wind-chill *table* [66]. The standard wind-chill *table* is based on a standard of calm of 3 mi / h (typical walking speed), rather than on the *true* standard of calm $\mathcal{V} = 0\,\mathrm{mi}\,/\,\mathrm{h}$ in *true* accordance with the standard wind-chill *formula* [Eq. (14)] that we adopt in this Sect. 4.2. Also, the recommended ranges of applicability of the standard wind-chill *table* are $-50\,^\circ\mathrm{F} < T \leq 50\,^\circ\mathrm{F}$ and $3\,\mathrm{mi}\,/\,\mathrm{h} < \mathcal{V} < 110\,\mathrm{mi}\,/\,\mathrm{h}$ [66]. But we base our calculations of $\mathcal{W}$ on the standard wind-chill *formula* [Eq. (14)], for which no limits on the range of applicability are stated for either T or $\mathcal{V}$ [66]. If there is a sufficiently strong wind on Neptune, then Eq. (14) yields a *cold* negative Kelvin *effective* wind-chill temperature $\mathcal{W}$.

A physical interpretation is this: In order to produce the same chilling effect as air at temperature $T = 72\,\mathrm{K} = -201\,^\circ\mathrm{C} = -330\,^\circ\mathrm{F}$ at the 1 bar level on Neptune [65] with a 50 mi / h wind [65,66], calm air would have to be at temperature $\mathcal{W} = -22\,\mathrm{K} = -296\,^\circ\mathrm{C} = -500\,^\circ\mathrm{F}$ — colder than absolute zero, sub-0 K. [The 0.1 bar level on Neptune is colder, but as noted in the first paragraph of this Sect. 4.2, Eq. (14) is likely more accurate if applied at the 1 bar level on Neptune.] The average thermal translational kinetic energy of the air molecules would have to be negative, and hence their average thermal speed imaginary. Our physical interpretation assumes that this super-cold, or hyper-cold, sub-0 K air of our imagination remains an ideal gas, for which the restricted definition of temperature as twice

[6] (Re: Entries [66] and [67], Refs. [66] and [67]) An online brochure accessible at Ref. [66] provides more information. Reference [67] augments Ref. [66] with still more information, including references and a few alternative formulas for wind-chill temperature $\mathcal{W}$. (In Australia the wind-chill temperature $\mathcal{W}$ is dubbed as the apparent temperature $\mathcal{AT}$.) In this Sect. 4.2. we always calculate $\mathcal{W}$ based on the formula employed by the U. S. A. National Weather Service [Eq. (14)].

the average thermal kinetic energy $\langle E_{\rm kin}\rangle$ per molecular translational degree of freedom divided by Boltzmann's constant, i.e., $T = 2\langle E_{\rm kin}\rangle / k$ [68–81], is valid — and is extrapolated as remaining valid even for negative values of $\langle E_{\rm kin}\rangle$ and T. A necessary, but probably not sufficient, property of our hypothetical super-cold, or hyper-cold, air molecules is that they exert no attractive forces, however weak, on each other, so that they could never condense into a liquid or solid. (This could obtain, at least for all practical purposes, if the average distance between real air molecules is more than a few orders of magnitude larger than typical molecular sizes of $\sim 10^{-10}$ m to $\sim 10^{-9}$ m — but then of course the density would be much lower than the $0.45\,{\rm kg}\,/\,{\rm m}^3$ obtaining at the 1 bar $\approx$ 1 atm level on Neptune.) Our physical interpretation seems limited to this restricted definition of temperature. There seems to be no obvious way of extending our physical interpretation of cold negative *effective* (wind-chill) Kelvin temperature $\mathcal{W}$ in terms of the most general definition of *true* Kelvin temperature, i.e., $T = (\partial E/\partial S)_{V,N}$ [68–81]. Even for *true* (not effective) *non*negative Kelvin temperatures, the restricted definition of temperature $T = 2\langle E_{\rm kin}\rangle / k$ [68–81] is valid if and only if, as is the case of ideal gases, $\langle E_{\rm kin}\rangle$ is directly proportional to T [68–81]. There exist excellent in-depth discussions of the concept of temperature, especially concerning the point that "Temperature is deeper than average kinetic energy." [78,79]. Nevertheless, although taking temperature as proportional to average thermal kinetic energy per degree of freedom is not the most general concept [78,79], it suffices to serve as one of the elements in an important derivation of Boltzmann's principle relating entropy and probability [80,81] and in an important generalization of the relation between entropy and heat [82].[7]

It has been argued [83], that Eq. (14) for wind-chill temperature $\mathcal{W}$ is only an approximation [83], that even as an approximation it is valid only at Earth-like or "human"

[7] (Re: Entries [71–77], Refs. [2], [3], [4], [48], [49], [76], and [77]) It is usually stated that the definition of temperature in terms of the Carnot efficiency of a reversible heat engine yields only a ratio of the two temperatures of the hot and cold reservoirs, not the one temperature of either reservoir considered individually, as does $T = (\partial E/\partial S)_{V,N}$, and as does even the more restricted $T = 2\langle E_{\rm kin}\rangle / k$ for the special case of ideal-gas reservoirs. To obtain actual values of temperature by this method rather than just the ratio of two temperatures, it is usually stated that the temperature of at least one of the two reservoirs must be ascertained by other means, most typically by allowing one reservoir to attain thermodynamic equilibrium with water at its triple point. This is discussed in Entries [71–75]. Thermodynamic equilibrium with water at its triple point is likewise employed to fix the temperature scale of ideal-gas thermometers, as discussed in Entry [75]. However, Refs. [76] and [77] describe how this requirement for a water-triple-point (or any other heat reservoir) is overcome, at least in principle even if not in practice, by employing a sequence of ideal, reversible, Carnot engines, the cold reservoir for engine $\mathbb{N}$ serving as the hot reservoir for engine $\mathbb{N}+1$, with the heat input for engine $\mathbb{N}+1$ to that for engine $\mathbb{N}$ being in a fixed ratio r ($0 < r < 1$), and with each engine doing an equal amount of work. Then the last engine in the sequence *must* have a cold reservoir at $T = 0\,{\rm K}$, thus dispensing with the requirement of a water-triple-point or other standardizing heat reservoir. Of course, this considers only the Second-Law aspect of the problem; according to the unattainability formulation of the third law, especially in its strongest mode, a cold reservoir at *precisely* $T = 0\,{\rm K}$ is impossible. But perhaps a cold reservoir at T *arbitrarily close to* 0 K or even *sufficiently close to* 0 K suffices to thus dispense, even if not perfectly then at least for all practical purposes, with the requirement for a water-triple-point or other standardizing heat reservoir. [One point concerning Sect. 58 of Ref. [77]: Consider, as in Sect. 58 of Ref. [77], a heat engine operating between a heat source at positive Kelvin temperature and a heat sink at *true* (*not* merely effective) *cold* negative Kelvin temperature *if* true (not merely effective) *cold* negative Kelvin temperatures could exist — *if* the Kelvin temperature scale *could* be linearly extrapolated downwards through $T = 0\,{\rm K}$ to *true* (not merely effective) negative values. Contrary to what is stated in Sect. 58 of Ref. [77], such a heat engine, *if* it could exist, would *not* discard more heat to its heat sink than it received from its heat source, thereby violating the First Law of Thermodynamics (conservation of energy). It would discard *negative* heat, i.e., it would *extract* heat, from its heat sink — in addition to extracting heat from its heat source as does a standard heat engine. Its work output (neglecting friction and other irreversibilities) would equal the heat it extracts from its heat source *plus* the heat it extracts from its heat sink. Thus its efficiency as usually defined = (work output) ÷ (heat extracted from heat source *alone*) > 100%, consistent with the First Law of Thermodynamics, but of course *in*consistent with the second and third laws.

http://dx.doi.org/10.5772/61641

temperatures [83]. Thus, even though Eq. (14) is likely more accurate if applied at the 1 bar level on Neptune than at the 0.1 bar level on Neptune, we cannot be sure of its accuracy even at the 1 bar level on Neptune [83]. Moreover, it has also been argued [83] that $\mathcal{W}$ is more correctly expressed as W / m^2 of heat loss flux rather than as the temperature of calm air that would have the same chilling effect as moving air — wind — at speed $\mathcal{V}$ [83]. Indeed, many if not most national weather services *do* express $\mathcal{W}$ as W / m^2 of heat loss flux rather than as the temperature of calm air that would have the same chilling effect. (The national weather services of the U. S. A. [66], Canada [67], and Australia [67] employ wind-chill formulas for the temperature of calm air that would have the same chilling effect.) But *if* wind chill *is* expressed as the temperature of calm air that would have the same chilling effect [66,67], *then* irrespective of the equation for $\mathcal{W}$ that even if not exactly correct is at least a good approximation at the 1 bar level on Neptune, be that Eq. (14) or otherwise, it seems inescapable that in order to produce the same chilling effect as a sufficiently strong wind at sufficiently cold but still positive Kelvin temperatures, calm air *must* be colder than 0 K. Thus it seems inescapable that the *effective* (wind-chill) Kelvin temperature $\mathcal{W}$ must then be colder than 0 K even if no *actual* temperature can be colder than 0 K.

This would obtain even more strongly for a helium atmosphere, which remains gaseous at a pressure of 1 bar at Kelvin temperature T which, while still positive, is nevertheless much colder than the value $T = 72\,\mathrm{K} = -201\,^\circ\mathrm{C} = -330\,^\circ\mathrm{F}$ obtaining at the 1 bar level on Neptune or even than the value $T = 55\,\mathrm{K} = -218\,^\circ\mathrm{C} = -361\,^\circ\mathrm{F}$ obtaining at the 0.1 bar level on Neptune [65,84,85].[8] While recognizing the caveats discussed in the immediately preceding paragraph, nevertheless for illustrative purposes and for argument's sake let us assume that the standard wind chill *formula* [Eq. (14)] retains at least approximate validity for gaseous helium at a pressure of 1 bar. At a pressure of 1 bar, the common isotope of naturally-occurring helium, ${}_2\mathrm{He}^4$, is gaseous at $T = 5\,\mathrm{K} = -268\,^\circ\mathrm{C} = -450\,^\circ\mathrm{F}$, and the rare isotope of naturally-occurring helium (which fortunately can be produced artificially [84]), ${}_2\mathrm{He}^3$, is gaseous at $T = 4\,\mathrm{K} = -269\,^\circ\mathrm{C} = -452\,^\circ\mathrm{F}$ [84,85]. Again taking $\mathcal{V} = 50$ mi / h, for $T = 5\,\mathrm{K} = -268\,^\circ\mathrm{C} = -451\,^\circ\mathrm{F}$ Eq. (14) yields $\mathcal{W} = -118\,\mathrm{K} = -391\,^\circ\mathrm{C} = -671\,^\circ\mathrm{F}$, and for $T = 4\,\mathrm{K} = -269\,^\circ\mathrm{C} = -452\,^\circ\mathrm{F}$ Eq. (14) yields $\mathcal{W} = -119\,\mathrm{K} = -392\,^\circ\mathrm{C} = -674\,^\circ\mathrm{F}$.

4.3. Limits of the possible

How impossible is the super-cold, or hyper-cold, sub-0 K air of our imagination — but with a true as opposed to merely effective sub-0 K temperature? It is (at the very least, almost) certainly physically impossible, but, at least *prima facie*, it seems not to be logically impossible. The physically impossible at least does not exist and possibly even cannot exist in physical reality, but can exist in the imagination and hence in virtual reality (imagination displayed via a computer). The logically impossible cannot exist — rather than merely does not exist — not only in physical reality, but to boot not even in the imagination and hence not even in virtual reality.

A Euclidean (planar) right triangle that violates the Pythagorean Theorem is not merely physically impossible but logically impossible. Such a triangle cannot exist — rather than

[8] (Re: Entry [85], Ref. [17]) In Fig. 14-19 (a) on p. 381 of Ref. [17], the normal boiling point of ${}_2\mathrm{He}^3$ is incorrectly shown at a pressure of approximately 1.3 atm instead of at 1 atm.

merely does not exist — not only in physical reality, but to boot not even in the imagination: it cannot even be imagined; it cannot exist even in virtual reality.

By contrast, for example, a violation of the first law of thermodynamics (conservation of energy) [86,87] is (at least so far as is known [86,87]) physically but not logically impossible — perpetual motion of the first kind can at least be *imagined*; it *can* exist at least in *virtual* reality [86,87]. Indeed even concerning the *physical* impossibility (or possibility?) of violation of energy conservation, we should note that energy conservation has never been rigorously proven in general relativity, and that there have been serious proposals for its possible violation at cosmological distance and time scales [86,87]. But: Any proposed violation of energy conservation should address the difficulty posed by Noether's Theorem [88]. According to Noether's Theorem [88], nonconservation of energy implies that the time-invariance of the fundamental laws of physics must be broken (and vice versa). There isn't much wiggle room — even small changes in the (at least apparent) fine-tuning of at least some of the laws of physics would render life (at least carbon-based life as we know it on Earth) impossible [89–92]. Energy, even free energy or equivalently negentropy, is far from being the only requirement for life. But could Noether's Theorem be satisfied by considering nascent energy to be a new boundary (initial) condition upon the Universe, thereby preserving the time-invariance of the laws of physics? For example, consider the following thought experiment: What if a mass m subject to local gravitational acceleration g could spontaneously rise through a height Δy to the ceiling — not spontaneously get cooler and rise to the ceiling (on demand rather than via unpredictable and uncontrollable fluctuation), thereby violating the Second Law of Thermodynamics, but just spontaneously rise to the ceiling, thereby violating the First Law of Thermodynamics (energy conservation)? Could the nascent gravitational potential energy $mg\Delta y$ simply be a new initial condition upon the Universe, leaving the time-invariance of the laws of physics intact? Might Noether's Theorem accept payment in the cheap currency of boundary (initial) conditions instead of the expensive currency of the laws of physics, and hence not pose any difficulty? [Note: Proposals such as those cited for genuine creation of nascent energy [86,87] should not be confused with proposals for creation of positive mass-energy at the expense of negative energy, typically at the expense of negative gravitational energy [93–101], but in some versions of the steady-state theory [102–107] at the expense of a negative-energy creation field (the C field) [105–107]. (There are difficulties associated with the C field [106,107].] The former proposals [86,87] but not the latter ones [93–107] contravene the First Law of Thermodynamics (conservation of energy).]

Thus since knowledge is imperfect and incomplete, perhaps one should not *a priori* rule out any nonzero probability, however remote, that a logically possible phenomenon might also be physically possible [86,87]. Hence the "Insofar as is known" in the first sentence of Sect. 4.2, the "(at the very least, almost)" in the second sentence of the first paragraph of Sect. 4.3, and the "(at least so far as is known [86,87])" in the first sentence of the third paragraph of Sect. 4.3. Unlike our Pythagorean-Theorem-violating Euclidean (planar) right triangle but like a violation of energy conservation [86,87], our super-cold, or hyper-cold, sub-0 K air can at least be *imagined*.

http://dx.doi.org/10.5772/61641

5. Brief concluding remarks

Hopefully our considerations of and related to absolute zero 0 K have been helpful. In Sect. 2.2, we showed that in principle 0 K can be *attained* at the expense of only a finite, typically small, cost of work via standard TSRR, in Sect. 2.5. at the expense of an even smaller cost of high-temperature heat via absorption TSRR, and in Sect. 3.1. at the expense of a small cost of work via CSRR, employing weighing or a Stern-Gerlach apparatus. (Recall from the sixth and seventh paragraphs of Sect. 3.2. that the specific QCR method discussed in Sect. 3 of Ref. [1] employs first CSRR and then ERR [1]: we employed the first, Stern-Gerlach-apparatus CSRR, step thereof in the sixth paragraph of Sect. 3.2.) In the standard and absorption TSRR cases, the unattainability formulation of the Third Law of Thermodynamics of thermodynamics does not require infinite expenditure of work and heat, respectively to attain 0 K, but forbids the expenditure, respectively, of the required small cost of work and even smaller cost of heat. But in the CSRR cases, it does *not*, even in its strongest mode, forbid the expenditure of the required small cost of work.

But there are also the difficulties of *maintaining* 0 K and of *verifying* [22] that 0 K has even been *attained*, which we discussed in Sect. 2.3, the last paragraphs of Sects. 2.4. and 2.5, and Sect. 3. *Perfectly maintaining* 0 K for more than infinitessimal time requires perfect insulation [23], and *perfectly verifying* [22] that 0 K has even been *attained* requires infinite time. Even given perfect insulation [23] (recall Sect. 2.3.) and hence that 0 K can be perfectly *maintained*, the unattainability formulation of the Third Law of Thermodynamics in its *strongest mode*, which forbids attainment of 0 K by *any means* whatsoever, seems inviolable with respect to *perfect* verification [22] that 0 K has been *attained*, because of the infinite-time requirement imposed by the energy-time uncertainty principle. But if we do not insist on *exactly* perfect verification [22] and are willing to accept verification that is perfect *for all practical purposes*, then to this extent the unattainability formulation of the third law even in its strongest mode *is* challenged. The limitation to "for all practical purposes" is further imposed because as per Sect. 2.3. insulation [23] can be perfect only for all practical purposes. At least in principle and possibly also in practice, CSRR and QCR [1] seem superior to standard TSRR or even absorption TSRR in effecting the challenge to the unattainability formulation of the Third Law of Thermodynamics in its strongest mode, albeit for all practical purposes and not with exact perfection.

Hopefully also our considerations in Sect. 4. of negative Kelvin temperatures, both true ones hotter than ∞ K and effective ones colder than 0 K, have been helpful.

6. Appendix: A few fine points concerning the Third Law of Thermodynamics

It is generally stated that the Nernst formulation of the Third Law of Thermodynamics, according to which all entropy changes vanish at 0 K, and the unattainability formulation thereof, according to which 0 K is unattainable in a finite number of finite operations, are equivalent. But we should note that there are dissensions to this viewpoint [108–113].[9]

[9] (Re: Entry [109], Ref. [109] Footnote 5 on p. 494 of Ref. [109] concerns "a residual inequivalence" between the Nernst heat theorem and unattainability principle, with the former construed as more fundamental.

Also, in considering the discreteness of energy eigenstates required by quantum mechanics in any system constrained within a fixed finite volume, or even within an unfixed but always finite volume for example corresponding to maintenance of constant pressure, we did not mention the role of quantum-mechanical Bose-Einstein symmetry or Fermi-Dirac antisymmetry requirements on the allowed wave functions [114]. The gaps between energy eigenstates at very low temperatures in light of these requirements can be much larger than would be the case in the absence of these requirements [114]. For a typical laboratory-type macroscopic system, the energy gap ΔE between the ground and first excited state is $\Delta E \sim 10^{-20}\,\mathrm{K}\,k - 10^{-19}\,\mathrm{K}\,k$ [114,115]. Yet the entropy and heat capacity of a typical laboratory-type macroscopic system is typically already only a very small fraction of the value predicted by classical (as opposed to quantum) statistical mechanics at $T \sim 10\,\mathrm{K}$ [114,115]. It has been noted that at $T \sim 10\,\mathrm{K}$ the energy *per particle* in a typical laboratory-type macroscopic system is $\sim \Delta E \sim 10^{-20}\,\mathrm{K}\,k - 10^{-19}\,\mathrm{K}\,k$ [114,115]. But because the characteristic temperatures of quantum statistical mechanics, for example, the Debye, Fermi-Dirac, and Bose-Einstein temperatures [116], are independent of the size of a system [116], this is a fortuitous result owing to the typical sizes of laboratory-type macroscopic systems [114,115].

A third formulation of the Third Law of Thermodynamics has also been stated [117], according to which the zero of entropy with a system in its ground energy eigenstate (assumed nondegenerate), is as unattainable as 0 K itself [117].[10] This hird formulation of the Third Law of Thermodynamics has been stated with respect to *thermo*dynamics. But, in fact, it ultimately obtains owing to the *pure* (quantum) dynamics of the energy-time uncertainty principle, and with respect to fixing a system *exactly* into *any* of its energy eigenstates in general (not just specifically its ground state), degenerate or not. The energy-time uncertainty principle requires infinite time to *exactly* — with *strictly zero* uncertainty — fix the energy of any system into *any* of its energy eigenstates in general (not just specifically its ground state), degenerate or not.

Acknowledgments

I gratefully acknowledge Dr. Marlan O. Scully for insightful and thoughtful discussions and communications concerning Ref. [1]. I also am very grateful to Dr. Donald H. Kobe for expressing interest in and providing helpful remarks concerning Ref. [1], as well as for acquainting me with Refs. [80–82]. Their contributions helped to inspire the ideas for this chapter. Additionally, I am indebted to a reviewer for the American Journal of Physics for valuable insights concerning an earlier version of this manuscript that was not accepted, and I am grateful to Dr. Daniel P. Sheehan for helpful comments and suggestions concerning earlier versions of this manuscript. I am thankful to Dr. Paolo Grigolini for very helpful and thoughtful considerations concerning both earlier versions and the most recent versions of this manuscript in Special Problems courses. I thank Dr. S. Mort Zimmerman for engaging in general scientific discussions over many years, and both Dan Zimmerman and Dr. Kurt W. Hess and for brief yet helpful discussions concerning this chapter and for engaging in general

[10] (Re: Entries [17] and [117], Refs. [17] and [117]) The detailed discussions concerning the third law of thermodynamics in Ref. [117] are largely deleted in Ref. [17], which provides only a brief mention of the third law on p. 217. Reference [117] dubs the Nernst formulation of the Third Law of Thermodynamics as the Nernst-Simon formulation thereof. Reference [17] does not render Ref. [117] obsolete, because Ref. [117] discusses aspects not discussed in Ref. [17], and vice versa.

scientific discussions at times. I also thank Dr. Iva Simcic, Publishing Process Manager, for much very helpful advice in preparing this chapter and for much extra time to prepare it, and Technical Support at MacKichan Software for their very helpful advice concerning Scientific WorkPlace 5.5.

Author details

Jack Denur

Address all correspondence to: jackdenur@my.unt.edu

Electric & Gas Technology, Inc., Rowlett, Texas, USA

References

[1] Scully MO, Aharonov Y, Kapale KT, Tannor DJ, Süssmann G, Walther H. Sharpening accepted thermodynamic wisdom via quantum control: or cooling to an internal temperature of zero by external coherent control fields without spontaneous emission. Journal of Modern Optics. 2002; **49**: 2297–2307. DOI: 10.1080/0950034021000011392

[2] Callen HB. Thermodynamics. New York: John Wiley & Sons; 1960, p. 27, Chap. 10 (especially Sect. 10.4), and Sect. 11.2.

[3] Callen HB. Thermodynamics and an Introduction to Thermostatistics. 2nd ed. New York: John Wiley & Sons; 1985, p. 30, Chap. 11 (especially Sect. 11-2), and Sect. 12-2.

[4] King, AL. Thermophysics. San Francisco: W. H. Freeman; 1962, Sects. 6.5 and 7.3, and p. 280.

[5] Guggenheim EA. Thermodynamics: An Advanced Treatment for Chemists and Physicists. 7th ed. Amsterdam: North-Holland; 1985, Sects. 1.66, 2.17, 3.53, 3.57–3.59, and 11.13–11.17 (especially 11.17).

[6] Kosloff R, Amikam L. Quantum Refrigerators and the III-Law of Thermodynamics. 12th Joint European Thermodynamics Conference; 1–5 July 2013; Brescia.

[7] Kosloff R, Amikam L. Quantum Heat Engines and Refrigerators: Continuous Devices. Ann. Rev. Phys. Chem. 2014; **65**: 365–393. DOI: 10.1146/annurev-physchem-040513-103724

[8] Allahverdyan AE, Hovhannisyan K, Mahler G. Optimal refrigerator. Phys. Rev. E. 2010; **81**. 051129-1–12.

[9] Reference [2], Chap. 4, especially Sects. 4.4– 4.7, and most especially Sect. 4-7.

[10] Reference [3], Chap. 4, especially Sects. 4.5–4-6, and most especially Sect. 4-6.

[11] Reference [4], pp. 57–82, especially Sects. 4.8, 5.3, 6.1, and 6.2.

[12] Reference [2], p. 75.

[13] Reference [3], pp. 115–116.

[14] Reference [2], Sects. 10.2 and 11.7.

[15] Reference [3], Chap. 11 (especially Sect. 11-2), Sects. 15-2, 15-3, 16-7, 16-8, 18-4, and 18-6.

[16] Reference [3], Sects. 15-3 and 16-2.

[17] Zemansky MW, Dittman RH. Heat and Thermodynamics. 7th ed. Boston: McGraw-Hill; 1997, Sects. 10.2 and 10.8.

[18] Wark K Jr. Thermodynamics. 6th ed. Boston: McGraw-Hill; 1999, Sects. 3-8 and 12-4.

[19] Reference [4], pp. 36, 237–240, and 316–326.

[20] Bohren CF, Albrecht BA. Atmospheric Thermodynamics. New York: Oxford University Press; 1998, pp. 335–366.

[21] Reference [4], Sect. 20.2.

[22] Reference [3], Sect. 11-3, especially the third sentence.

[23] Baturina TI, Vinokur VM. Superinsulator-Superconductor Duality in Two Dimensions. Ann. Phys. 2013; **331**: 236–257. DOI: 10.1016/j.aop.2012.12.007

[24] Reference [4], Sects. 5.8 and 6.1.

[25] Einstein A, Szilárd L. Refrigeration. U. S. A. Patent No. 1,781,541; 1930.

[26] Reference [2], p. 190.

[27] Reference [2], Sect. 10.4.

[28] Private communication, reviewer for the American Journal of Physics. 2010.

[29] Cohen BL. A simple treatment of potential barrier penetration. Am. J. Phys. 1965; **33**: 97–98. DOI: 10.1119/1.1971334

[30] Gomer R. Field Emission and Field Ionization. Cambridge, Mass: Harvard University Press; 1961, pp. 6–11, 66–70, and 181–183. The application of the position-momentum uncertainty principle per se is discussed on pp. 6–8.

[31] Hagmann MJ. Transit time for quantum tunneling. Solid State Comm. 1992; **82**: 867–870. DOI: 10.1016/0038-1098(92)90710-Q

[32] Hagmann MJ. Quantum tunneling times: A new solution compared to 12 other methods. Int. J. Quant. Chem. Suppl. 1992; **26**: 299–309. DOI: 10.1002/qua.560440826

[33] Hagmann MJ. Distribution of times for barrier traversal caused by energy fluctuations. J. App. Phys. 1993; **74**: 7302–7305. DOI: 10.1063/1.354995

[34] Hagmann MJ. Effects of the finite duration of quantum tunneling in laser-assisted scanning tunneling microscopy. Int. J. Quant. Chem. Suppl. 1994; **28**: 271–282. DOI: 10.1002/qua.560520829

[35] Hagmann MJ. Reduced effects of laser illumination due to the finite duration of quantum tunneling. J. Vac. Sci. Technol. B. 1994; **12**: 3191–3195. DOI: 10.1116/1.587498

[36] Landau LD, Lifshitz EM. Statistical Physics. 3rd ed. Oxford: Pergamon; 1980 (2005 Printing), Sect. 110.

[37] Landau LD, Lifshitz EM. Quantum Mechanics. 2nd revised ed. Oxford: Butterworth-Heinemann; 1991 (2005 Printing), Sects. 1, 16, and 44.

[38] Kobe DH, Aguilera-Navarro VC. Derivation of the energy-time uncertainty relation. Phys. Rev. A. 1994; **50**: 933–938. DOI: 10.1103/PhysRevA.50.933

[39] Kobe DH, Iwamoto H, Goto M, Aguilera-Navarro VC. Derivation of the energy-time uncertainty relation. Phys. Rev. A. 2001; **64**: Article ID: 022104: 8 pages. DOI: 10.1103/PhysRevA.64.022104

[40] Dodonov VV, Dodonov AV. Energy-time and frequency-time uncertainty relations: exact inequalities. Physica Scripta. 2015; **90**: Article ID: 074049: 22 pages. DOI: 10.1088/0031-8949/90/7/074049

[41] Denur J. The energy-time uncertainty principle and quantum phenomena. Am. J. Phys. 2010; **78**: 1132–1145. DOI: 10.1119/1.3133084

[42] Hill TL. Statistical Mechanics: Principles and Selected Applications. Original copyright: New York: McGraw-Hill; 1956. Unabridged and unaltered republication: New York: Dover; 1987. Chapters 1 and 2; especially Sects. 4, 5, 10, 11, and 13.

[43] Tolman RC. The Principles of Statistical Mechanics. Original copyright: Oxford, U. K.: Clarendon Press; 1938. Unabridged and unaltered republication: New York: Dover; 1979. Sections 21–25, pp. 344–345, and Sects. 84 and 98(e).

[44] Bohm D. Quantum Theory. Original copyright: Englewood Cliffs, N. J.: Prentice-Hall; 1951. Unaltered and unabridged republication: New York: Dover; 1989. Chapter 23.

[45] Reference [44], Chap. 8 (especially Sects. 8.22 and 8.23) and Chap. 22 (especially Sects. 22.2–22.3).

[46] Eisner RM. Fundamentals of Modern Physics. New York: John Wiley & Sons; 1961, Sect. 7.7.

[47] Reference [46], Sect. 2.5.

[48] Baeirlein R. Atoms and Information Theory. San Francisco: W. H. Freeman; 1971, Sect. 10.6, and Problem 10.8 of Chap. 10 on pp. 369–371.

[49] Baeirlein R. Thermal Physics. Cambridge, U. K.: Cambridge University Press; 1999, Chap. 6.

[50] Reference [46], Sects. 3.6–3.7, especially Sect. 3.7.

[51] Reference [46], Sects. 6.3.

[52] Brillouin L. Relativity Reexamined. New York: Academic; 1970, Sects. 1.1 and 3.3–3.4.

[53] Reference [46], Sect. 7.5.

[54] Sheehan DP, editor. Quantum Limits to the Second Law, AIP Conference Proceedings Volume 643. Melville, N. Y.: American Institute of Physics; 2002.

[55] Nukulov AV, Sheehan DP, editors. Special Issue: Quantum Limits to the Second Law of Thermodynamics. Entropy. March 2004: Vol. 6, Issue 1.

[56] Čápek V, Sheehan DP. Challenges to the Second Law of Thermodynamics: Theory and Experiment. Dordrecht, The Netherlands: Springer; 2005.

[57] Sheehan DP, editor. Special Issue: The Second Law of Thermodynamics: Foundations and Status. Found. Phys. December 2007: Vol. 37, Issue 12.

[58] Sheehan DP, editor. Second Law of Thermodynamics: Status and Challenges, AIP Conference Proceedings Volume 1411. Melville, N. Y.: American Institute of Physics; 2011.

[59] Reference [4], Sect. 18.4 and relevant references cited therein, especially our Ref. [60] immediately following.

[60] Ramsey NF. Thermodynamics and Statistical Mechanics at Negative Absolute Temperatures. Phys. Rev. 1956: **103**; 20–28. DOI: 10.1103/PhysRev.103.20

[61] Reference [48], Sect. 11.5.

[62] Reference [49], Sect. 14.6.

[63] Atkins P. Four Laws that Drive the Universe. Oxford, U. K.: Oxford University Press; 2007, Chap. 5, especially p. 118.

[64] Reference [20], Sects. 2.2–2.3, especially Fig. 2.7 on p. 56.

[65] Williams DR. Neptune Fact Sheet [Internet]. 2015. Available from: nssdc.gsfc.nasa.gov/planetary/factsheet/neptunefact.html. [Accessed: 2015-12-05]

[66] NWS Windchill Chart [Internet]. 2001. Available from: www.nws.noaa.gov/om/winter/windchill.shtml [Accessed: 2015-12-05]

[67] See the Wikipedia article entitled "Wind chill." [Internet]. 2015. Available from: www.wikipedia.org [Accessed: 2015-12-05]

[68] Reference [2], Sects. 2.1–2.6 and 4.10–4.11.

[69] Reference [3], Sects. 2-1–2-6 and 4-8.

[70] Reference [4], Chaps. 1 and 2, and Sects. 9.1 and 11.1–11.6.

[71] Reference [2], Sect. 4.9.

[72] Reference [3], Sect. 4-8.

[73] Reference [4], Sects. 6.4–6.6.

[74] Reference [48], Sect. 4.2.

[75] Reference [49], pp. 82–86.

[76] Beep NR. Meteorological Thermodynamics and Atmospheric Statics. In Berry FA, Bollay E, Beers NR, editors. Handbook of Meteorology. New York: McGraw-Hill; 1945, pp. 320–325 and 332–337, especially pp. 321–323 and 335–337.

[77] Faries VM. Applied Thermodynamics. Revised Edition. New York: Macmillan; 1949, Sect. 58.

[78] Reference [48], Sects. 4.3 and 7.5.

[79] Reference [49], Sects. 1.1–1.2, pp. 177–178, and Sect. 14.7.

[80] Campisi M, Kobe DH. Derivation of Boltzmann Principle. arXiv:0911.2070v1 [cond-mat.stat-mech].11 Nov 2009. 9 pages. [Internet]. 2009. Available from: xxx.lanl.gov/abs/0911.2070 or arxiv.org/abs/0911.2070 [Accessed: 2015-12-05]

[81] Campisi M, Kobe DH. Derivation of the Boltzmann principle. Am. J. Phys. 2010; **78**: 608–615. DOI: 10.1119/1.3298372

[82] Campisi M., Bagci GB. Tsallis ensemble as an exact orthode. Phys. Lett. A 2007; **362**: 11–15. DOI: 10.1016/j.physleta.2006.09.081

[83] Private communication, reviewer for the American Journal of Physics. 2010.

[84] Reference [4], Sect. 19.1.

[85] Reference [17], Sect. 14.6.

[86] Sheehan DP, Kriss VG: Energy Emission by Quantum Systems in an Expanding FRW Metric. arXiv:astro-ph/0411299v1. 11 Nov 2004. 12 pages. [Internet]. 2004. Available from: xxx.lanl.gov/abs/astro-ph/0411299 or arxiv.org/abs/astro-ph/0411299 [Accessed: 2015-12-05]

[87] Harrison ER. Mining Energy in an Expanding Universe. Astroph. J. 1995: **446**; 63–66. DOI: 10.1086/175767

[88] Reference [3], Chap. 21, especially Sects. 21-2–21-4.

[89] Vilenkin A. Many Worlds in One. New York: Hill and Wang; 2007, pp. 132–139 and 144–151. DOI: 10.1063/1.2743129

[90] Tegmark M. Our Mathematical Universe. New York: Vintage Books; 2015, pp. 138–145, 150, and 352–365.

[91] Carr B, editor. Universe or Multiverse? Cambridge, U. K.: Cambridge University Press; 2007. See especially Chaps. 1, 2, 3, 5, 22, 23, and 25.

[92] Rindler W. Relativity: Special, General, and Cosmological. 2nd ed. New York: Oxford University Press; 2006, Sect. 18.6 and references cited therein.

[93] Reference [89], pp. 11–12 and 183–186.

[94] Davies PCW. The Physics of Time Asymmetry. Berkeley: University of California Press; 1977, pp. 190–191 and 199–200.

[95] Tryon EP. Is the Universe a Vacuum Fluctuation? Nature 1973; **246**: 396–397. DOI: 10.1038/246396a0

[96] Vilenkin A. Many Worlds in One: The Search for Other Universes. New York: Hall and Wang; 2007. DOI: 10.1063/1.2743129, pp. 183–186.

[97] Filippenko A, Pasachoff JM. A Universe From Nothing. 2 pages. [Internet]. 2001. Available from: https:www.astrosociety.org/publications/a-universe-from-nothing [Accessed: 2015-12-05]

[98] Potter F, Jargodski C. Mad About Modern Physics. Hoboken, NJ; 2005, Question 224, "The Total Energy," on p. 115, and Answer 224, "The Total Energy," along with cited references, on pp. 275–276.

[99] Berman MS. On the Zero-Energy Universe. Int. J. Theor. Phys. 2009; **48**: 3278–3286. DOI: 10.1007/s10773-009-0125-8

[100] Berman MS, Trevisan LA. On Creation of Universe Out of Nothing. Int. J. Modern Phys. B. 2010; **19**: 1309–1313. DOI: 10.1142/S021827181007342

[101] Penrose R. The Road to Reality: A Complete Guide to the Laws of the Universe. New York: Alfred A. Knopf; 2005, Sects. 19.7–19.8 (especially Sect. 19.8 and most especially pp. 468–469). See also references cited therein, especially those cited in Note 19.17 on p. 470, and Notes for Sects. 19.7–19.8 on pp. 469–470.

[102] Vilenkin A. Many Worlds in One: The Search for Other Universes. New York: Hall and Wang; 2007. DOI: 10.1063/1.2743129, pp. 27–28, 37, and 170–171.

[103] Rindler W. Relativity: Special, General, and Cosmological. 2nd ed. New York: Oxford University Press; 2006, pp. 359, 368–369, 387–388, and 411.

[104] Bradley W. Carroll BW, Ostlie DA. An Introduction to Modern Astrophysics. 2nd ed. San Francisco: Pearson Addison Wesley; 2007, pp. 1163–1167.

[105] Reference [94], Sect. 7.2 and references cited therein.

[106] Reference [94], the last sentence on p. 187 and the reference cited therein.

[107] Davies PCW. Is the Universe Transparent or Opaque? J. Phys. A: Math. Gen. 1972; **5**: 1722–1737. DOI: 10.1088/0305-4470/5/12/012

[108] Baierlein R. Thermal Physics. Cambridge, U. K.: Cambridge University Press; 1999, Chap. 14, especially Sects.14.4–14.5, 14.7, 14.9, and relevant references cited under "Further Reading" on pp. 352–353.

[109] Berry RS, Rice SA, Ross J. Physical Chemistry. 2nd ed. New York: Oxford University Press; 2000, Chap. 18, especially Sect. 18.2.

[110] Wheeler JC. Nonequivalence of the Nernst-Simon and unattainability statements of the third law of thermodynamics. Phys. Rev. A 1991; **43**: 5289–5295. DOI: 10.1103/PhysRevA.43.5289

[111] Wheeler JC. Addendum to "Nonequivalence of the Nernst-Simon and untenability statements of the third law of thermodynamics." Phys. Rev. A 1992; **45**: 2637–2640. DOI: 10.1103/PhysRevA.45.2637

[112] Landsberg PT. A comment on Nernst's theorem. J. Phys. A. Math. Gen. 1989; **22**:139–141. DOI: 10.1088/0305-4470/22/1/021

[113] Landsberg PT. Answer to Question #34. What is the third law of thermodynamics trying to tell us? Am. J. Phys. 1997; **65**: 269–270. DOI: 10.1119/1.18483

[114] Garrod C. Statistical Mechanics and Thermodynamics. New York; Oxford University Press; 1995, Sect. 5.16 (especially the discussion concerning Axiom 5), Sect. 5.18, Problem 5.19 on p. 137, and, as auxiliary material, Appendix A.10 and Exercise 5.11 on pp. 441–443.

http://dx.doi.org/10.5772/61641

[115] Reference [108], Sect.14.4.

[116] Reference [108], Sects. 6.5, 9.1, 9.4, and 14.4.

[117] Zemansky MW, Dittman RH. Heat and Thermodynamics. 6th ed. New York: McGraw-Hill; 1981, Sect. 19-6 and Problems for Chap. 19 on pp. 520–521, especially Problems 19-2–19-4.

Chapter 11

Foundation of Equilibrium Statistical Mechanics Based on Generalized Entropy

A.S. Parvan

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/60997

Abstract

The general mathematical formulation of the equilibrium statistical mechanics based on the generalized statistical entropy for the first and second thermodynamic potentials was given. The Tsallis and Boltzmann-Gibbs statistical entropies in the canonical and microcanonical ensembles were investigated as an example. It was shown that the statistical mechanics based on the Tsallis statistical entropy satisfies the requirements of equilibrium thermodynamics in the thermodynamic limit if the entropic index z=1/(q-1) is an extensive variable of state of the system.

Keywords: Equilibrium statistical mechanics, Tsallis nonextensive statistics

1. Introduction

In modern physics, there exist alternative theories for the equilibrium statistical mechanics [1, 2] based on the generalized statistical entropy [3-12]. They are compatible with the second part of the second law of thermodynamics, i.e., the maximum entropy principle [13-14], which leads to uncertainty in the definition of the statistical entropy and consequently the equilibrium probability density functions. This means that the equilibrium statistical mechanics is in a crisis. Thus, the requirements of the equilibrium thermodynamics shall have an exclusive role in selection of the right theory for the equilibrium statistical mechanics. The main difficulty in foundation of the statistical mechanics based on the generalized statistical entropy, i.e., the deformed Boltzmann-Gibbs entropy, is the problem of its connection with the equilibrium thermodynamics. The proof of the zero law of thermodynamics and the principle of additivity

in general serves as a primarily problem. The equilibrium thermodynamics is a phenomenological theory defined on the class of homogeneous functions of the zero and first order [15].

The formalism of the statistical mechanics agrees with the requirements of the equilibrium thermodynamics if the thermodynamic potential, which contains all information about the physical system, in the thermodynamic limit is a homogeneous function of the first order with respect to the extensive variables of state of the system [14, 6-7]. It was proved that for the Tsallis and Boltzmann-Gibbs statistics [6, 7], the Renyi statistics [10], and the incomplete nonextensive statistics [12], this property of thermodynamic potential provides the zeroth law of thermodynamics, the principle of additivity, the Euler theorem, and the Gibbs-Duhem relation if the entropic index z is an extensive variable of state. The scaling properties of the entropic index z and its relation to the thermodynamic limit for the Tsallis statistics were first discussed in the papers [16, 17].

The aims of this study are to establish the connection between the Tsallis statistics, i.e., the statistical mechanics based on the Tsallis statistical entropy, and the equilibrium thermodynamics and to prove the zero law of thermodynamics.

The structure of the chapter is as follows. In Section 2, we review the basic postulates of the equilibrium thermodynamics. The equilibrium statistical mechanics based on generalized entropy is formulated in a general form in Section 3. In Section 4, we describe the Tsallis statistics and analyze its possible connection with the equilibrium thermodynamics. The main conclusions are summarized in the final section.

2. Equilibrium thermodynamics

2.1. Thermodynamic potentials

In the equilibrium thermodynamics, the physical properties of the system are fully identified by the fundamental thermodynamic potential $f = f(x_1, \ldots, x_n)$ as a real-valued function of n real variables, which are called the variables of state. The macroscopic state of the system is fixed by the set of independent variables of state $x = (x_1, \ldots, x_n)$. Each variable of state x_i, which is related to the certain thermodynamic quantity, describes some individual property of the system. The first and the second partial derivatives of the thermodynamic potential with respect to the variables of state define the thermodynamic quantities (observables) of the system, which describe other individual properties of this system. The first differential and the first partial derivatives of the fundamental thermodynamic potential with respect to the variables of state can be written as

$$df = \sum_{i=1}^{n} u_i dx_i, \qquad u_i = \frac{\partial f}{\partial x_i}, \tag{1}$$

http://dx.doi.org/10.5772/60997

where the vector $u=(u_1, \ldots, u_n)$. Equation (1) is the fundamental equation of thermodynamics and expresses the first law of thermodynamics. The second differential of the fundamental thermodynamic potential is written as a quadratic form

$$d^2 f = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} dx_i dx_j, \quad a_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}, \tag{2}$$

where a_{ij} is the element of the symmetric matrix A of the dimension $(n \times n)$. The symmetry conditions for the matrix elements, $a_{ij}=a_{ji}$, lead to the equalities (Maxwell relations)

$$\frac{\partial u_j}{\partial x_i} = \frac{\partial u_i}{\partial x_j} \qquad (i, j = 1, \ldots, n). \tag{3}$$

If the function $f(x_1, \ldots, x_n)$ is convex (concave) on $s \le n$ variables of state, then the quadratic form (Eq. (2)) in s variables is positive definite (negative definite). The quadratic form (Eq. (2)) in s variables for which $a_{ij}=a_{ji}$ is positive definite (negative definite) if [18]

$$a_{11}\xi > 0, \ \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \xi^2 > 0, \ \ldots, \ \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1s} \\ a_{21} & a_{22} & \cdots & a_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ a_{s1} & a_{s2} & \cdots & a_{ss} \end{vmatrix} \xi^s > 0 \tag{4}$$

for every nonzero vector x, where $\xi=1$ for the positive definite quadratic form and $\xi=-1$ for the negative definite quadratic form. Note that the fundamental thermodynamic potential f, the set of variables of state x, and the vector u constitute the complete set of $2n+1$ variables, which completely define the given thermodynamic system.

The first thermodynamic potential $g=g(y)$ is a function of a new set of independent variables of state $y=(u_1, \ldots, u_m, x_{m+1}, \ldots, x_n)$, which is obtained by the Legendre transform from the fundamental thermodynamic potential $f(x_1, \ldots, x_n)$ changing $m \le n$ variables of state $(x_1, \ldots, x_m)$ with their conjugate variables $(u_1, \ldots, u_m)$. The set of unknown variables $x_1, \ldots, x_m$ is a solution of a system of m differential equations [19]:

$$\frac{\partial f}{\partial x_i} = u_i \qquad (i = 1, \ldots, m). \tag{5}$$

Solving this system of equations, we obtain m functions of the variables of state,

$$x_i = x_i(u_1,\ldots,u_m,x_{m+1},\ldots,x_n) \quad (i=1,\ldots,m). \tag{6}$$

Substituting Eq. (6) into the fundamental thermodynamic potential f and using the Legendre transform, we obtain [19]

$$g = f - \sum_{i=1}^{m} \frac{\partial f}{\partial x_i} x_i = f - \sum_{i=1}^{m} u_i x_i. \tag{7}$$

This Legendre transform is always well defined when the fundamental thermodynamic potential $f(x_1, \ldots, x_n)$ is a convex function of the variables $(x_1, \ldots, x_m)$, i.e., the quadratic form $\sum_{i,j=1}^{m} a_{ij} d x_i d x_j$ is positive definite [19]. To obtain this, it is necessary and sufficient to satisfy the relations (4) for $s=m$ [18].

The first differential and the first partial derivatives of the first thermodynamic potential g can be written as

$$\begin{aligned} dg &= \sum_{i=1}^{n} v_i dy_i, \qquad v_i = \frac{\partial g}{\partial y_i}, \\ v_i &= \frac{\partial g}{\partial u_i} = -x_i \qquad (i=1,\ldots,m), \\ v_i &= \frac{\partial g}{\partial x_i} = u_i \qquad (i=m+1,\ldots,n). \end{aligned} \tag{8}$$

The second differential and the second partial derivatives of the first thermodynamic potential g are

$$d^2 g = \sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij} dy_i dy_j, \quad b_{ij} = \frac{\partial^2 g}{\partial y_i \partial y_j}. \tag{9}$$

The symmetry conditions for the matrix elements, $b_{ij}=b_{ji}$, impose the following equalities

$$\begin{aligned} \frac{\partial x_j}{\partial u_i} &= \frac{\partial x_i}{\partial u_j} \qquad (i,j=1,\ldots,m), \\ \frac{\partial u_j}{\partial u_i} &= -\frac{\partial x_i}{\partial x_j} \qquad (i=1,\ldots,m, j=m+1,\ldots,n), \\ \frac{\partial u_j}{\partial x_i} &= \frac{\partial u_i}{\partial x_j} \qquad (i,j=m+1,\ldots,n). \end{aligned} \tag{10}$$

http://dx.doi.org/10.5772/60997

If the function $g(u_1, \ldots, u_m, x_{m+1}, \ldots, x_n)$ is convex (concave) on $s \le n$ variables of state, then the quadratic form (Eq. (9)) in s variables is positive definite (negative definite). The quadratic form (Eq. (9)) in s variables for which $b_{ij} = b_{ji}$ is positive definite (negative definite) if [18]

$$b_{11}\xi > 0, \begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} \xi^2 > 0, \ldots, \begin{vmatrix} b_{11} & b_{12} & \ldots & b_{1s} \\ b_{21} & b_{22} & \ldots & b_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ b_{s1} & b_{s2} & \ldots & b_{ss} \end{vmatrix} \xi^s > 0 \tag{11}$$

for every nonzero vector y, where $\xi = 1$ for the positive definite quadratic form and $\xi = -1$ for the negative definite quadratic form.

The Legendre transform (Eq. (7)) is involutive [19], i.e., if under the Legendre transformation f is taken to g, then the Legendre transform of g will again be f. The fundamental thermodynamic potential $f(x_1, \ldots, x_n)$ can be obtained from the first thermodynamic potential $g(u_1, \ldots, u_m, x_{m+1}, \ldots, x_n)$ by the Legendre back-transformation

$$f = g - \sum_{i=1}^{m} \frac{\partial g}{\partial u_i} u_i = g + \sum_{i=1}^{m} x_i u_i, \tag{12}$$

where m functions $u_i = u_i(x_1, \ldots, x_n)$, $i = 1, \ldots, m$ are the solutions of the system of m differential equations

$$\frac{\partial g}{\partial u_i} = -x_i \quad (i = 1, \ldots, m). \tag{13}$$

This Legendre transform is well defined when the function $g(u_1, \ldots, u_m, x_{m+1}, \ldots, x_n)$ is a convex function of the variables $(u_1, \ldots, u_m)$, i.e., the quadratic form $\sum_{i,j=1}^{m} b_{ij} d u_i d u_j$ is positive definite [19]. To obtain this, it is necessary and sufficient to satisfy the relations (Eq. (11)) for $s = m$ [18].

The second thermodynamic potential $h = h(r)$ is obtained from the fundamental thermodynamic potential $f = f(x_1, \ldots, x_n)$ by expressing the variable x_k through the set of independent variables $r = (x_1, \ldots, x_{k-1}, f, x_{k+1}, \ldots, x_n)$:

$$h = x_k(x_1, \ldots, x_{k-1}, f, x_{k+1}, \ldots, x_n), \tag{14}$$

where, now, x_k is the second thermodynamic potential and f is a variable of state. The condition of independence of the variables of state r can be written as

$$\frac{\partial f}{\partial x_i}+\frac{\partial f}{\partial h}\frac{\partial h}{\partial x_i}=0. \tag{15}$$

Then the first differential and the first partial derivatives of the second thermodynamic potential h can be written as

$$dh=\sum_{i=1}^{n} w_i dr_i, \quad w_i=\frac{\partial h}{\partial r_i}, \tag{16}$$

$$\begin{aligned} w_i &= \frac{\partial x_k}{\partial f}=\frac{1}{u_k} \qquad (i=k), \\ w_i &= \frac{\partial x_k}{\partial x_i}=-\frac{u_i}{u_k} \quad (i=1,\ldots,n, i\neq k). \end{aligned} \tag{17}$$

The second differential and the second partial derivatives of the second thermodynamic potential h can be written as

$$d^2h=\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} dr_i dr_j, \qquad c_{ij}=\frac{\partial^2 h}{\partial r_i \partial r_j}, \tag{18}$$

$$c_{ii}=\frac{\partial^2 x_k}{\partial f^2}=-\frac{1}{u_k^2}\frac{\partial u_k}{\partial f} \quad (i=k). \tag{19}$$

The symmetry conditions for the matrix elements, $c_{ij}=c_{ji}$, impose the following equalities

$$\begin{aligned} \frac{u_i}{u_k^2}\frac{\partial u_k}{\partial f}-\frac{1}{u_k}\frac{\partial u_i}{\partial f} &= -\frac{1}{u_k^2}\frac{\partial u_k}{\partial x_i} && (i\neq k), \\ \frac{u_j}{u_k^2}\frac{\partial u_k}{\partial x_i}-\frac{1}{u_k}\frac{\partial u_j}{\partial x_i} &= \frac{u_i}{u_k^2}\frac{\partial u_k}{\partial x_j}-\frac{1}{u_k}\frac{\partial u_i}{\partial x_j} && (i,j\neq k). \end{aligned} \tag{20}$$

2.2. Principle of additivity

In the equilibrium thermodynamics, all thermodynamic quantities belong to the class of homogeneous functions of zero and first order, which imposes the additional constraints on the thermodynamic system. The homogeneous function of k th order satisfies the relation [14, 15, 20]

http://dx.doi.org/10.5772/60997

$$f\left(\lambda x_1,\ldots,\lambda x_m,x_{m+1},\ldots,x_n\right)=\lambda^k f\left(x_1,\ldots,x_m,x_{m+1},\ldots,x_n\right) \tag{21}$$

and the Euler theorem[1]

$$\sum_{i=1}^{m}\frac{\partial f}{\partial x_i}x_i=kf, \tag{22}$$

where $(x_1, \ldots, x_m)$ are extensive variables and $(x_{m+1}, \ldots, x_n)$ are intensive variables. Note that the function f is extensive if $k=1$, and it is intensive if $k=0$.

Let us divide the system into two subsystems: system(1+2) = system(1) + system(2).

Then m -extensive and $n-m$ -intensive variables of state satisfy the additivity relations [6, 15]

$$\begin{aligned} x_i^{1+2}&=x_i^1+x_i^2 \qquad \left(i=1,\ldots,m\right),\\ x_i^{1+2}&=x_i^1=x_i^2 \qquad \left(i=m+1,\ldots,n\right). \end{aligned} \tag{23}$$

The homogeneous function of the first degree ($k=1$), which is extensive, is an additive function of the first order [15]

$$f^{1+2}\left(x_1^{1+2},\ldots,x_n^{1+2}\right)=f^1\left(x_1^1,\ldots,x_n^1\right)+f^2\left(x_1^2,\ldots,x_n^2\right) \tag{24}$$

and the homogeneous function of the zero degree ($k=0$), which is intensive, is an additive function of zero order [15]

$$f^{1+2}\left(x_1^{1+2},\ldots,x_n^{1+2}\right)=f^1\left(x_1^1,\ldots,x_n^1\right)=f^2\left(x_1^2,\ldots,x_n^2\right). \tag{25}$$

Note that the zero law of thermodynamics is expressed by Eqs. (21) and (25) when the temperature T is a function or the second equation of Eq. (23) when temperature T is a variable of state.

3. Equilibrium statistical mechanics

In comparison with the equilibrium thermodynamics, the system in the equilibrium statistical mechanics is described by two additional elements: the microstates of the system and the

1 In this subsection, the symbol f denotes any function not only the fundamental thermodynamic potential.

probabilities of these microstates. As in the equilibrium thermodynamics, the macrostates of the system are fixed by the set of independent variables of state. The thermodynamic potential is a universal function that depends not only on the macroscopic state variables of the system but also on the microstates of the system and their probabilities. The extensive thermodynamic quantities are calculated as averages over the ensemble of microstates. However, the intensive thermodynamic quantities are defined in terms of the first derivatives of the thermodynamic potential with respect to the extensive variables of state.

Let us formulate the main statements of the equilibrium statistical mechanics. Let the thermodynamic potential be a function $Y = Y(p_1, \ldots, p_W; X^1, \ldots, X^n)$ of W -independent variables $(p_1, \ldots, p_W)$ and n variables of state $(X^1, \ldots, X^n)$. All arguments of the function Y are independent.[2]

The first thermodynamic potential $Y = g(p_1, \ldots, p_W; u^1, \ldots, u^m, x^{m+1}, \ldots, x^n)$ is a function of m intensive variables of state $X^j = u^j$ $(j = 1, \ldots, m)$ conjugated to the variables $(x^1, \ldots, x^m)$ and $n - m$ extensive variables of state $X^j = x^j$ $(j = m + 1, \ldots, n)$. The first thermodynamic potential Y is related to the fundamental thermodynamic potential $f = f(p_1, \ldots, p_W; x^1, \ldots, x^n)$ by the Legendre transform (7) [19]

$$Y = f - \sum_{j=1}^{m} u^j x^j, \qquad u^j = \frac{\partial f}{\partial x^j}. \tag{26}$$

Here and in the following, the first thermodynamic potential will be considered only for the statistical ensembles for which $x^1 = S$, $X^1 = u^1 = T$, and $f = E$, where S is the entropy, T is the temperature, and E is the energy.

The second thermodynamic potential $Y = h = x^k(p_1, \ldots, p_W; x^1, \ldots, x^{k-1}, f, x^{k+1}, \ldots, x^n)$ is a function of $n - 1$ variables of state $X^j = x^j$ $(j = 1, \ldots, n, j \neq k)$ and one variable $X^k = f$ for $1 \leq k \leq n$. In the following, the second thermodynamic potential will be associated only with the microcanonical ensemble $(k = 1)$ for which $Y = x^1 = S$ and $X^1 = f = E$.

Let Y_i and $x_i^1, \ldots, x_i^n$ be the values of the dynamical extensive variables Y and $x^1, \ldots, x^n$, respectively, in the i th microscopic state of the system. Moreover, let us impose an additional constraint on the variables $(p_1, \ldots, p_W)$ [18],

$$\varphi\left(p_1, \ldots, p_W; X^1, \ldots, X^n\right) = \sum_i \delta_{X^{m+1}, X_i^{m+1}} \ldots \delta_{X^n, X_i^n} p_i - 1 = 0, \tag{27}$$

where $\delta_{x,x'}$ is the Kronecker delta for the integer x, x' and the Dirac delta function for the real x, x'. In Eq. (27), the variables $X_i^j = x_i^j$ $(j = m + 1, \ldots, n)$ are for the first thermodynamic potential,

2 In this section, the thermodynamic quantities are numbered by the index at the top. The index at the bottom of the variable denotes the microstate of the system.

http://dx.doi.org/10.5772/60997

and $m=0$, $X^1=f=E$, $X_i^1=f_i=E_i$, and $X_i^j=x_i^j$ ($j=2, \ldots, n$) are for the second thermodynamic potential.

The ensemble averages for the extensive dynamical variables A can be written as

$$A\left(p_1,\ldots,p_W;X^1,\ldots,X^n\right)=\sum_i \delta_{X^{m+1},X_i^{m+1}}\ldots\delta_{X^n,X_i^n}p_iA_i, \tag{28}$$

where A_i is the value of the variable A in the i th microscopic state of the system.

The first and the second thermodynamic potentials, which are the extensive functions of the variables of state, can also be written as (28)

$$Y(p_1,\ldots,p_W;X^1,\ldots,X^n)=\sum_i \delta_{X^{m+1},X_i^{m+1}}\ldots\delta_{X^n,X_i^n}p_iY_i, \tag{29}$$

$$Y_i=f_i-\sum_{j=1}^{m}u^jx_i^j \qquad \text{for } Y=g, \tag{30}$$

$$Y_i=x_i^1=S_i \qquad \text{for } Y=x^1=S, \tag{31}$$

where S_i and f_i are the values of the entropy S and the fundamental thermodynamic potential f, respectively, in the i th microstate of the system, which are both determined by Eq. (28).

In the equilibrium statistical mechanics, the unknown probabilities of microstates $\{p_i\}$ are found from the second part of the second law of thermodynamics, i.e., from the constrained extremum of the thermodynamic potential (Eq. (29)) as a function of the variables ($p_1, \ldots, p_W$) under the condition that the variables ($p_1, \ldots, p_W$) satisfy Eq. (27). Moreover, it is supposed that the value of the entropy in the i th microstate of the system is a function of the probability p_i of this microstate, i.e., $x_i^1=S_i=S_i(p_i)$. Then to determine the unknown probabilities $\{p_i\}$ at which the function Y attains the constrained local extrema, the method of Lagrange multipliers [18] can be used

$$\Phi\left(p_1,\ldots,p_W;X\right)=Y\left(p_1,\ldots,p_W;X\right)-\lambda\varphi\left(p_1,\ldots,p_W;X\right), \tag{32}$$

$$\frac{\partial\Phi(p_1,\ldots,p_W;X)}{\partial p_i}=0 \qquad (i=1,\ldots,W), \tag{33}$$

where λ is an arbitrary constant and the vector $X=(X^1, \ldots, X^n)$. Substituting Eqs. (27) and (29) into Eq. (32), we obtain

$$Y_i + p_i \frac{\partial Y_i}{\partial p_i} - \lambda = 0 \quad \left(i = 1,\ldots,W\right). \tag{34}$$

Substituting Eq. (30) into Eq. (34) and using Eq. (27), we obtain the probabilities related to the first thermodynamic potential

$$p_i = \psi\left(\frac{1}{u^1}\left(\Lambda - f_i + \sum_{j=2}^{m} u^j x_i^j\right)\right), \tag{35}$$

$$\sum_i \delta_{x^{m+1},x_i^{m+1}} \ldots \delta_{x^n,x_i^n} \psi\left(\frac{1}{u^1}\left(\Lambda - f_i + \sum_{j=2}^{m} u^j x_i^j\right)\right) = 1, \tag{36}$$

where $\Lambda \equiv \phi(\lambda) = \Lambda(u^1, \ldots, u^m, x^{m+1}, \ldots, x^n)$ is the solution of Eq. (36) and $\psi(x)$ is a function related to the given function $x_i^1 = S_i(p_i)$.

Substituting Eq. (31) into Eq. (34) and using Eq. (27), we obtain the probabilities related to the second thermodynamic potential

$$p_i = \frac{1}{W\left(f, x^2, \ldots, x^n\right)}, \tag{37}$$

$$W\left(f, x^2, \ldots, x^n\right) = \sum_i \delta_{f,f_i} \delta_{x^2,x_i^2} \ldots \delta_{x^n,x_i^n}, \tag{38}$$

where $p_i = \psi(\lambda)$ is a constant the same for all microstates of the system. Note that in these derivations, the conditions $\partial f_i / \partial p_i = 0$, $\partial u^j / \partial p_i = 0$ ($j = 1, \ldots, m$), and $\partial x_i^j / \partial p_i = 0$ ($j = 2, \ldots, m$) were used.

Let us consider the first thermodynamic potential. Substituting Eq. (35) into Eq. (26) and using Eq. (36), we obtain the expression for the first thermodynamic potential as

$$\begin{aligned} Y\left(u^1,\ldots,u^m,x^{m+1},\ldots,x^n\right) &= f\left(u^1,\ldots,u^m,x^{m+1},\ldots,x^n\right) \\ &\quad - \sum_{j=1}^{m} u^j x^j \left(u^1,\ldots,u^m,x^{m+1},\ldots,x^n\right). \end{aligned} \tag{39}$$

Then the partial derivatives of the first thermodynamic potential (Eq. (39)) with respect to the variables of state can be written as

http://dx.doi.org/10.5772/60997

$$x^k = -\frac{\partial Y}{\partial u^k} = \sum_i \delta_{x^{m+1},x_i^{m+1}} \ldots \delta_{x^n,x_i^n} p_i x_i^k \qquad \left(k = 1,\ldots,m\right), \tag{40}$$

$$\begin{aligned} u^k = \frac{\partial Y}{\partial x^k} &= \sum_i \delta_{x^{m+1},x_i^{m+1}} \ldots \delta_{x^n,x_i^n} p_i \frac{\partial f_i}{\partial x^k} \\ &+ \sum_i \delta_{x^{m+1},x_i^{m+1}} \ldots \delta_{x^n,x_i^n} \left(\frac{\partial \delta_{x^k,x_i^k}}{\partial x^k} \right) p_i Y_i \qquad \left(k = m+1,\ldots,n\right). \end{aligned} \tag{41}$$

Here we used the conditions $\partial f_i / \partial u^k = 0$, $\partial x_i^j / \partial u^k = 0$ ($j = 2, \ldots, n$) and Eqs. (34) and (36).

The fundamental thermodynamic potential can be written as

$$f = Y - \sum_{j=1}^{m} u^j \frac{\partial Y}{\partial u^j} = \sum_i \delta_{x^{m+1},x_i^{m+1}} \ldots \delta_{x^n,x_i^n} p_i f_i. \tag{42}$$

Let us consider the second thermodynamic potential. Substituting Eqs. (37), (38), and (31) into Eq. (29), we obtain the expression for the second thermodynamic potential

$$Y\left(f, x^2, \ldots, x^n\right) = S_i\left(p_i\right), \tag{43}$$

where p_i is determined from Eq. (37) and $Y = S$. The partial derivatives (Eq. (17)) of the second thermodynamic potential (Eq. (43)) with respect to the variables of state can be written as

$$\frac{1}{u^1} = \frac{\partial Y}{\partial f} = -\gamma \frac{\partial \ln W}{\partial f}, \qquad \gamma \equiv p_i \frac{\partial S_i\left(p_i\right)}{\partial p_i}, \tag{44}$$

$$\frac{u^j}{u^1} = -\frac{\partial Y}{\partial x^j} = \gamma \frac{\partial \ln W}{\partial x^j} - \left. \frac{\partial S_i}{\partial x^j} \right|_{p_i = const} \qquad \left(j = 2,\ldots,n\right), \tag{45}$$

where W is determined from Eq. (38).

Finally, it should be mentioned that the equilibrium statistical mechanics is thermodynamically self-consistent if the statistical variables ($x^1, \ldots, x^n$), the potentials ($f, g, \ldots$), and the variables ($u^1, \ldots, u^n$) are homogeneous variables of the first- or zero-order satisfying Eqs. (21)-(25).

4. Tsallis statistical mechanics

The Tsallis statistical mechanics is based on the generalized entropy which is a function of the entropic parameter q and probing probabilities p_i [3, 4]:

$$S = \sum_i p_i S_i, \qquad S_i = k_B z\left(1 - p_i^{1/z}\right), \tag{46}$$

where $z = 1/(q-1)$, k_B is the Boltzmann constant, and $q \in \mathbb{R}$ is a real parameter, $0 < q < \infty$. In the limit $q \to 1$ ($z \to \pm\infty$), the entropy (Eq. (46)) recovers the usual Boltzmann-Gibbs entropy $S = \sum_i p_i S_i$, where $S_i = -k_B \ln p_i$ [3]. Note that throughout the work, we use the system of natural units $\hbar = c = k_B = 1$.

4.1. Canonical ensemble

The thermodynamic potential of the canonical ensemble, the Helmholtz free energy, is the first thermodynamic potential $g = F$, which is a function of the variables of state $u^1 = T$, $x^2 = V$, $x^3 = N$, and $x^4 = z$. It is obtained from the fundamental thermodynamic potential $f = E$ (the energy) by the Legendre transform (Eq. (7)), exchanging the variable of state $x^1 = S$ of the fundamental thermodynamic potential with its conjugate variable $u^1 = T$. In the canonical ensemble, the first partial derivatives (Eq. (1)) of the fundamental thermodynamic potential are defined as $u^2 = -p$, $u^3 = \mu$, and $u^4 = -\Xi$. The entropy (Eq. (46)) for the Tsallis and Boltzmann-Gibbs statistics in the canonical ensemble can be rewritten as

$$S = \sum_i \delta_{V,V_i} \delta_{N,N_i} \delta_{z,z_i} p_i S_i, \tag{47}$$

$$S_i = z_i (1 - p_i^{1/z_i}) \qquad \text{for } |z| < \infty, \tag{48}$$

$$S_i = -\ln p_i \qquad \text{for } |z| = \infty. \tag{49}$$

The first thermodynamic potential (Eqs. (26) and (29)), $Y = F$, for the Tsallis and Boltzmann-Gibbs statistics can be rewritten as

$$F = E - TS = \sum_i \delta_{V,V_i} \delta_{N,N_i} \delta_{z,z_i} p_i F_i, \qquad F_i \equiv E_i - TS_i, \tag{50}$$

$$F_i = E_i - T z_i \left(1 - p_i^{1/z_i}\right) \qquad \text{for } |z| < \infty, \tag{51}$$

http://dx.doi.org/10.5772/60997

$$F_i = E_i + T \ln p_i \qquad \text{for } |z| = \infty. \tag{52}$$

Here the constraint (Eq. (27)) in the canonical ensemble is in the form

$$\varphi = \sum_i \delta_{V,V_i} \delta_{N,N_i} \delta_{z,z_i} p_i - 1 = 0. \tag{53}$$

Applying the method of Lagrange multipliers (Eqs. (32)-(34)) with the Lagrange function $\Phi = F - \lambda\varphi$ to Eqs. (50)-(53), we can write down Eqs. (34) and (35) for the Tsallis and Boltzmann-Gibbs statistics immediately as

$$F_i + p_i \frac{\partial F_i}{\partial p_i} - \lambda = 0 \tag{54}$$

and [7]

$$p_i = \left[1 + \frac{1}{z_i + 1} \frac{\Lambda - E_i}{T} \right]^{z_i} \quad \text{for } |z| < \infty, \tag{55}$$

$$p_i = e^{\frac{\Lambda - E_i}{T}} \qquad \text{for } |z| = \infty, \tag{56}$$

where $\Lambda \equiv \lambda - T$ and $\partial E_i / \partial p_i = 0$. Then the constraint (Eq. (53)) for the probabilities (Eqs. (55) and (56)) of the Tsallis and Boltzmann-Gibbs statistics can be written as [7]

$$\sum_i \delta_{V,V_i} \delta_{N,N_i} \delta_{z,z_i} \left[1 + \frac{1}{z_i + 1} \frac{\Lambda - E_i}{T} \right]^{z_i} = 1 \quad \text{for } |z| < \infty, \tag{57}$$

$$\sum_i \delta_{V,V_i} \delta_{N,N_i} \delta_{z,z_i} e^{\frac{\Lambda - E_i}{T}} = 1 \qquad \text{for } |z| = \infty, \tag{58}$$

where $\Lambda = \Lambda(T, V, N, z)$ is the solution of Eq. (57) for the Tsallis statistics and $\Lambda = -T \ln Z_G$ is the solution of Eq. (58) for the Boltzmann-Gibbs statistics, where $Z_G = \sum_i \delta_{V,V_i} \delta_{N,N_i} \delta_{z,z_i} e^{-E_i/T}$ is the partition function. Substitution of the probabilities given in Eqs. (55) and (56) into Eqs. (50)-(53) gives the Helmholtz free energy [7]

$$F = \frac{z}{z+1}\left(\Lambda + \frac{E}{z}\right) \qquad \text{for } |z| < \infty, \tag{59}$$

$$F = \Lambda = -T \ln Z_G \qquad \text{for } |z| = \infty, \tag{60}$$

where E is the energy (Eq. (42)), which can be written in terms of the canonical ensemble as

$$E = F - T\frac{\partial F}{\partial T} = \sum_i \delta_{V,V_i}\delta_{N,N_i}\delta_{z,z_i} p_i E_i. \tag{61}$$

Making use of Eqs. (40), (50), and (54), we can write the entropy of the system as

$$S = -\frac{\partial F}{\partial T} = \sum_i \delta_{V,V_i}\delta_{N,N_i}\delta_{z,z_i} p_i S_i. \tag{62}$$

Here we have used the conditions that the derivative of the constraint (Eq. (53)) with respect to T is zero, $\partial E_i / \partial T = 0$, $\partial E_i / \partial p_i = 0$, and $\partial T / \partial p_i = 0$. Substituting Eqs. (48), (49), (55), and (56) into Eq. (62) and using Eqs. (57) and (58), we obtain [7]

$$S = -\frac{z}{z+1}\frac{\Lambda - E}{T} \qquad \text{for } |z| < \infty, \tag{63}$$

$$S = -\frac{\Lambda - E}{T} \qquad \text{for } |z| = \infty. \tag{64}$$

Using Eqs. (41), (50), and (54), we obtain the pressure, $u^2 = -p$, and the chemical potential, $u^3 = \mu$:

$$-p = \frac{\partial F}{\partial V} = \sum_i \delta_{V,V_i}\delta_{N,N_i}\delta_{z,z_i} p_i \frac{\partial E_i}{\partial V} + \sum_i \delta_{N,N_i}\delta_{z,z_i}\left(\frac{\partial \delta_{V,V_i}}{\partial V}\right) p_i F_i, \tag{65}$$

$$\mu = \frac{\partial F}{\partial N} = \sum_i \delta_{V,V_i}\delta_{N,N_i}\delta_{z,z_i} p_i \frac{\partial E_i}{\partial N} + \sum_i \delta_{V,V_i}\delta_{z,z_i}\left(\frac{\partial \delta_{N,N_i}}{\partial N}\right) p_i F_i. \tag{66}$$

Here we have used the conditions that the derivatives of the constraint (Eq. (53)) with respect to the variables of state N and V are zero, $\partial E_i / \partial p_i = 0$ and $\partial T / \partial p_i = 0$.

http://dx.doi.org/10.5772/60997

Substituting Eqs. (50) and(54) into Eq. (41), we obtain the variable Ξ :

$$-\Xi = \frac{\partial F}{\partial z} = -T\sum_i \delta_{V,V_i}\delta_{N,N_i}\delta_{z,z_i}p_i\left[1-p_i^{1/z_i}\left(1-\ln p_i^{1/z_i}\right)\right] + \sum_i \delta_{V,V_i}\delta_{N,N_i}\left(\frac{\partial \delta_{z,z_i}}{\partial z}\right)p_i F_i \qquad \text{for } |z|<\infty, \tag{67}$$

$$-\Xi = \frac{\partial F}{\partial z} = 0 \qquad \text{for } |z|=\infty, \tag{68}$$

where we have used the conditions that the derivative of the constraint (53) with respect to z is zero, $\partial E_i/\partial z=0$, $\partial E_i/\partial p_i=0$, and $\partial T/\partial p_i=0$.

Thus, from the results given in Eqs. (62) and (65)-(67), we see that the differential of the thermodynamic potential (Eq. (50)) satisfies [7, 15]

$$dF = -SdT - pdV + \mu dN - \Xi dz. \tag{69}$$

Using Eqs. (50) and (69), we obtain the fundamental equation of thermodynamics [7, 14, 15]

$$TdS = dE + pdV - \mu dN + \Xi dz. \tag{70}$$

To prove the homogeneity properties of the thermodynamic quantities and the Euler theorem for the Tsallis statistics in the canonical ensemble, we will consider, as an example, the exact analytical results for the nonrelativistic ideal gas.

4.1.1. Nonrelativistic ideal gas: canonical ensemble

Let us investigate the nonrelativistic ideal gas of identical particles governed by the classical Maxwell-Boltzmann statistics in the framework of the Tsallis and Boltzmann-Gibbs statistical mechanics.

It is convenient to obtain the exact results for the ideal gas in the Tsallis statistics by means of the integral representation for the Gamma function (see [9] and reference therein):

$$x^{-y} = \frac{1}{\Gamma(y)}\int_0^\infty dt t^{y-1}e^{-tx}, \qquad \mathrm{Re}\, x>0,\ \mathrm{Re}\, y>0, \tag{71}$$

$$x^{y-1} = \Gamma(y)\frac{i}{2\pi}\oint_C dt\left(-t\right)^{-y}e^{-tx}, \qquad \mathrm{Re}\, x>0,\ |y|<\infty. \tag{72}$$

Thus, solving Eqs. (57) and (58) for the ideal gas in the canonical ensemble, the norm function Λ is [7]

$$1+\frac{1}{z+1}\frac{\Lambda}{T}=\left[Z_G\frac{\Gamma\left(-z-3N/2\right)}{\left(-z-1\right)^{-3N/2}\Gamma\left(-z\right)}\right]^{-\frac{1}{z+3N/2}} \quad \text{for } z<-1, \tag{73}$$

$$1+\frac{1}{z+1}\frac{\Lambda}{T}=\left[Z_G\frac{\left(z+1\right)^{3N/2}\Gamma\left(z+1\right)}{\Gamma\left(z+1+3N/2\right)}\right]^{-\frac{1}{z+3N/2}} \quad \text{for } z>0, \tag{74}$$

$$Z_G=\frac{\left(gV\right)^N}{N!}\left(\frac{mT}{2\pi}\right)^{\frac{3}{2}N} \tag{75}$$

and $\Lambda=-T\ln Z_G$ for $|z|=\infty$. Here m is the particle mass, and Eq. (73) is restricted by the condition $-z-3N/2>0$.

The energy (Eq. (61)) and the thermodynamic potentials (Eqs. (59) and (60)) for the ideal gas in the canonical ensemble for the Tsallis and Boltzmann-Gibbs statistics can be written as [7]

$$E=\frac{3}{2}TN\eta,\quad \eta=\left(1+\frac{1}{z+1}\frac{\Lambda}{T}\right)\left(1+\frac{1}{z+1}\frac{3}{2}N\right)^{-1} \quad \text{for } |z|<\infty, \tag{76}$$

$$F=-T\left[z-\left(z+\frac{3}{2}N\right)\eta\right] \quad \text{for } |z|<\infty \tag{77}$$

and $E=3TN/2$ and $F=\Lambda=-T\ln Z_G$ for $|z|=\infty$.

The entropies (Eqs. (63) and (64)) and the pressure (Eq. (65)) for the ideal gas in the canonical ensemble for the Tsallis and Boltzmann-Gibbs statistics can be written as [7]

$$S=z\left(1-\eta\right) \quad \text{for } |z|<\infty, \tag{78}$$

$$p=\frac{N}{V}T\,\eta=\frac{2}{3}\frac{E}{V} \quad \text{for } |z|<\infty \tag{79}$$

and $S=\ln Z_G+3N/2$, $p=NT/V$ for $|z|=\infty$.

http://dx.doi.org/10.5772/60997

The chemical potential (Eq. (66)) and the variable (Eqs. (67) and (68)) for the ideal gas in the canonical ensemble for the Tsallis and Boltzmann-Gibbs statistics become [7]

$$\mu = -T\eta\left[\ln\left(gV\left(\frac{mT}{2\pi}\gamma_1(z+1)\left(1+\frac{1}{z+1}\frac{\Lambda}{T}\right)\right)^{3/2}\right)-\frac{3}{2}\frac{1}{z+1+3N/2}\right] + T\eta\left[\psi(N+1)+\frac{3}{2}\psi\left(a_1+\gamma_1\frac{3}{2}N\right)\right] \quad \text{for } |z|<\infty, \tag{80}$$

$$\Xi = T\left[1-\eta\left(1+\frac{z}{(z+1)^2}\frac{3}{2}N\right)\left(1+\frac{1}{z+1}\frac{3}{2}N\right)^{-1}\right] + T\eta\left[\ln\left(1+\frac{1}{z+1}\frac{\Lambda}{T}\right)+\psi(a_1)-\psi\left(a_1+\gamma_1\frac{3}{2}N\right)\right] \quad \text{for } |z|<\infty \tag{81}$$

and $\mu = -T\left[\ln\left(gV(mT/2\pi)^{3/2}\right)-\psi(N+1)\right]$, $\Xi = 0$ for $|z| = \infty$, where $\psi(y)$ is the psi-function, $a_1 = -z$, $\gamma_1 = -1$ for $z < -1$ and $a_1 = z+1$, $\gamma_1 = 1$ for $z > 0$.

4.1.2. Nonrelativistic ideal gas in the thermodynamic limit: canonical ensemble

Let us try to express the thermodynamic quantities of the nonrelativistic ideal gas directly in terms of the thermodynamic limit when the entropic parameter z is considered as an extensive variable of state

$$V \to \infty,\ N \to \infty,\ |z| \to \infty,\ v = \frac{V}{N} = const,\ \tilde{z} = \frac{z}{N} = const. \tag{82}$$

Note first that the canonical partition function (Eq. (75)) for the nonrelativistic ideal gas for the Boltzmann-Gibbs statistics can be rewritten as

$$Z_G = \tilde{Z}_G{}^N,\ \tilde{Z}_G \equiv gve\left(\frac{mT}{2\pi}\right)^{3/2}. \tag{83}$$

The norm functions (Eqs. (73) and (74)) for the ideal gas in the thermodynamic limit in the Tsallis and Boltzmann-Gibbs statistics can be rewritten as [7]

$$\Lambda = -TN\left[\tilde{z}-\left(\tilde{z}+\frac{3}{2}\right)\left(\tilde{Z}_G e^{3/2}\right)^{-\frac{1}{\tilde{z}+\frac{3}{2}}}\right] \quad \text{for } |\tilde{z}|<\infty, \tag{84}$$

$$\Lambda = -TN\ln\tilde{Z}_G \qquad \text{for } |\tilde{z}| = \infty, \tag{85}$$

where Eq. (84) is restricted by the conditions $\tilde{z} < -3/2$ and $\tilde{z} > 0$.

In the thermodynamic limit (Eq. (82)), the energy of the system (Eq. (76)) and the thermodynamic potential (Eq. (77)) for the ideal gas in the canonical ensemble for the Tsallis and Boltzmann-Gibbs statistics become [7]

$$E = \frac{3}{2}TN\left(\tilde{Z}_G e^{3/2}\right)^{-\frac{1}{\tilde{z}+\frac{3}{2}}} \qquad \text{for } |\tilde{z}| < \infty, \tag{86}$$

$$E = \frac{3}{2}TN \qquad \text{for } |\tilde{z}| = \infty \tag{87}$$

and [7]

$$F = \Lambda = -TN\left[\tilde{z} - \left(\tilde{z}+\frac{3}{2}\right)\left(\tilde{Z}_G e^{3/2}\right)^{-\frac{1}{\tilde{z}+\frac{3}{2}}}\right] \quad \text{for } |\tilde{z}| < \infty, \tag{88}$$

$$F = \Lambda = -TN\ln\tilde{Z}_G \qquad \text{for } |\tilde{z}| = \infty, \tag{89}$$

respectively. The entropy (78) and the pressure (79) for the ideal gas in the Tsallis and Boltzmann-Gibbs statistics in the thermodynamic limit can be written as [7]

$$S = N\tilde{z}\left[1 - \left(\tilde{Z}_G e^{3/2}\right)^{-\frac{1}{\tilde{z}+\frac{3}{2}}}\right] \qquad \text{for } |\tilde{z}| < \infty, \tag{90}$$

$$S = N\left(\ln\tilde{Z}_G + 3/2\right) \qquad \text{for } |\tilde{z}| = \infty \tag{91}$$

and [7]

$$p = \frac{T}{v}\left(\tilde{Z}_G e^{3/2}\right)^{-\frac{1}{\tilde{z}+\frac{3}{2}}} = \frac{2}{3}\frac{\varepsilon}{v} \qquad \text{for } |\tilde{z}| < \infty, \tag{92}$$

$$p = \frac{T}{v} = \frac{2}{3}\frac{\varepsilon}{v} \qquad \text{for } |\tilde{z}| = \infty, \tag{93}$$

http://dx.doi.org/10.5772/60997

where $\varepsilon = E/N$ is the specific energy given in Eqs. (86) and (87).

In the thermodynamic limit (82), the chemical potential (80) and the variable (81) for the ideal gas in the canonical ensemble for the Tsallis and Boltzmann-Gibbs statistics are [7]

$$\mu = T\left(\tilde{Z}_G e^{3/2}\right)^{-\frac{1}{\tilde{z}+\frac{3}{2}}}\left[\frac{5}{2}+\tilde{z}\ln\left(\tilde{Z}_G e^{3/2}\right)^{-\frac{1}{\tilde{z}+\frac{3}{2}}}\right] \quad \text{for } |\tilde{z}|<\infty, \tag{94}$$

$$\mu = T\left(1-\ln\tilde{Z}_G\right) \qquad \text{for } |\tilde{z}|=\infty \tag{95}$$

and [7]

$$\Xi = T\left[1-\left(\tilde{Z}_G e^{3/2}\right)^{-\frac{1}{\tilde{z}+\frac{3}{2}}}\left(1-\ln\left(\tilde{Z}_G e^{3/2}\right)^{-\frac{1}{\tilde{z}+\frac{3}{2}}}\right)\right] \quad \text{for } |\tilde{z}|<\infty, \tag{96}$$

$$\Xi = 0 \qquad \text{for } |\tilde{z}|=\infty. \tag{97}$$

Thus, from the results for the Tsallis statistics given in Eqs. (86), (90), (92), (94), and (96), we see that the Euler theorem (Eq. (22)) is satisfied [7]

$$TS = E + pV - \mu N + \Xi z. \tag{98}$$

Moreover, the thermodynamic quantities (86), (88), (90), (92), (94) and (96) satisfy the relation for the thermodynamic potential

$$F = E - TS = -pV + \mu N - \Xi z. \tag{99}$$

Next we shall verify that, when the entropic parameter z is an extensive variable of state in the thermodynamic limit, the ideal gas is in accordance with the principle of additivity. Suppose that the system is divided into two subsystems (1 and 2). Then the extensive variables of state of the canonical ensemble are additive

$$V^{1+2} = V^1 + V^2, \quad N^{1+2} = N^1 + N^2, \quad z^{1+2} = z^1 + z^2. \tag{100}$$

However, the temperature and the specific variables of state (Eq. (82)) are the same in each subsystem

$$T^{1+2} = T^1 = T^2, \quad v^{1+2} = v^1 = v^2, \quad \tilde{z}^{1+2} = \tilde{z}^1 = \tilde{z}^2. \tag{101}$$

Considering Eqs. (83), (100), and (101), we can verify that the Tsallis thermodynamic potential (Eq. (88)) and the entropy (Eq. (90)) of the canonical ensemble are homogeneous functions of the first order, i.e., $F(T, V, N, z)/N = F(T, v, \tilde{z})$ and $S(T, V, N, z)/N = S(T, v, \tilde{z})$, respectively, and they are additive (extensive)

$$F^{1+2}\left(T^{1+2}, V^{1+2}, N^{1+2}, z^{1+2}\right) = F^1\left(T^1, V^1, N^1, z^1\right) + F^2\left(T^2, V^2, N^2, z^2\right), \tag{102}$$

$$S^{1+2}\left(T^{1+2}, V^{1+2}, N^{1+2}, z^{1+2}\right) = S^1\left(T^1, V^1, N^1, z^1\right) + S^2\left(T^2, V^2, N^2, z^2\right). \tag{103}$$

The Tsallis pressure (Eq. (92)), the chemical potential (Eq. (94)), and the variable (Eq. (96)) are the homogeneous functions of the zero order, i.e., $p(T, V, N, z) = p(T, v, \tilde{z})$, $\mu(T, V, N, z) = \mu(T, v, \tilde{z})$, and $\Xi(T, V, N, z) = \Xi(T, v, \tilde{z})$, respectively, and they are the same in each subsystem (intensive)

$$p^{1+2}\left(T^{1+2}, V^{1+2}, N^{1+2}, z^{1+2}\right) = p^1\left(T^1, V^1, N^1, z^1\right) = p^2\left(T^2, V^2, N^2, z^2\right), \tag{104}$$

$$\mu^{1+2}\left(T^{1+2}, V^{1+2}, N^{1+2}, z^{1+2}\right) = \mu^1\left(T^1, V^1, N^1, z^1\right) = \mu^2\left(T^2, V^2, N^2, z^2\right), \tag{105}$$

$$\Xi^{1+2}\left(T^{1+2}, V^{1+2}, N^{1+2}, z^{1+2}\right) = \Xi^1\left(T^1, V^1, N^1, z^1\right) = \Xi^2\left(T^2, V^2, N^2, z^2\right). \tag{106}$$

Thus, the principle of additivity (Eqs. (21), (24), and (25)) is totally satisfied by the Tsallis statistics. Equations (101) and (103) prove the zero law of thermodynamics for the canonical ensemble.

4.2. Microcanonical ensemble

The thermodynamic potential of the microcanonical ensemble, the entropy, is the second thermodynamic potential $h = x^1 = S$ defined in Eq. (14), which is a function of the variables of state $f = E$, $x^2 = V$, $x^3 = N$ and $x^4 = z$. It is obtained from the fundamental thermodynamic potential f by exchanging the variable of state x^1 with variable f. In the microcanonical ensemble, the first partial derivatives of the fundamental thermodynamic potential (1) are defined as $u^1 = T$, $u^2 = -p$, $u^3 = \mu$, and $u^4 = -\Xi$.

The entropy S for the Tsallis and Boltzmann-Gibbs statistics in the microcanonical ensemble can be written as

http://dx.doi.org/10.5772/60997

$$S = \sum_i \delta_{E,E_i} \delta_{V,V_i} \delta_{N,N_i} \delta_{z,z_i} p_i S_i, \tag{107}$$

where S_i is defined in Eqs. (48) and (49). The set of probabilities $\{p_i\}$ is constrained by Eq. (27):

$$\varphi = \sum_i \delta_{E,E_i} \delta_{V,V_i} \delta_{N,N_i} \delta_{z,z_i} p_i - 1 = 0. \tag{108}$$

Applying the method of Lagrange multipliers (Eqs. (32)-(34)) with the Lagrange function $\Phi = S - \lambda\varphi$ to Eqs. (107), (108), (48), and (49), we can write down Eqs. (34), (37), and (38) for the Tsallis and Boltzmann-Gibbs statistics immediately as [6]

$$S_i + p_i \frac{\partial S_i}{\partial p_i} - \lambda = 0, \tag{109}$$

$$p_i = \frac{1}{W}, \tag{110}$$

$$W = \sum_i \delta_{E,E_i} \delta_{V,V_i} \delta_{N,N_i} \delta_{z,z_i}, \tag{111}$$

where $W = W(E, V, N)$ is the statistical weight for the Tsallis and Boltzmann-Gibbs statistics and $z_i = z$ for all microstates. Substituting Eqs. (110) and (111) into Eqs. (107), (48), and (49) and using Eq. (108), we can express the second thermodynamic potential (Eq. (43)) as [6]

$$S = z\left(1 - W^{-1/z}\right) \qquad \text{for } |z| < \infty, \tag{112}$$

$$S = \ln W \qquad \text{for } |z| = \infty. \tag{113}$$

Then the first derivative (Eq. (44)) of the thermodynamic potential S with respect to the variable of state E, i.e., the temperature T, can be rewritten as [6]

$$\frac{1}{T} = \frac{\partial S}{\partial E} = W^{-1/z} \frac{\partial \ln W}{\partial E} \qquad \text{for } |z| < \infty, \tag{114}$$

$$\frac{1}{T} = \frac{\partial S}{\partial E} = \frac{\partial \ln W}{\partial E} \qquad \text{for } |z| = \infty. \tag{115}$$

The first derivative (Eq. (45)) of the thermodynamic potential S with respect to the variable of state V, i.e., the pressure p, becomes [6]

$$p = T\frac{\partial S}{\partial V} = TW^{-1/z}\frac{\partial \ln W}{\partial V} \qquad \text{for } |z| < \infty, \tag{116}$$

$$p = T\frac{\partial S}{\partial V} = T\frac{\partial \ln W}{\partial V} \qquad \text{for } |z| = \infty. \tag{117}$$

The first derivative (Eq. (45)) of the thermodynamic potential S with respect to the variable of state N, i.e., the chemical potential μ, is [6]

$$\mu = -T\frac{\partial S}{\partial N} = -TW^{-1/z}\frac{\partial \ln W}{\partial N} \qquad \text{for } |z| < \infty, \tag{118}$$

$$\mu = -T\frac{\partial S}{\partial N} = -T\frac{\partial \ln W}{\partial N} \qquad \text{for } |z| = \infty. \tag{119}$$

The first derivative (Eq. (45)) of the thermodynamic potential S with respect to the variable of state z, i.e., the variable Ξ, can be rewritten as [6]

$$\Xi = T\frac{\partial S}{\partial z} = T\left[1 - W^{-1/z}\left(1 - \ln W^{-1/z}\right)\right] \qquad \text{for } |z| < \infty, \tag{120}$$

$$\Xi = 0 \qquad \text{for } |z| = \infty, \tag{121}$$

where $\partial W/\partial z = 0$. Then the differential of the thermodynamic potential (107) satisfies the fundamental equation of thermodynamics (70).

4.2.1. Nonrelativistic ideal gas: microcanonical ensemble

Let us consider the nonrelativistic ideal gas of N identical particles governed by the classical Maxwell-Boltzmann statistics in the framework of the Tsallis and Boltzmann-Gibbs statistics in the microcanonical ensemble. For this special model, the statistical weight (111) can be written as (see [6] and reference therein)

$$W = \frac{(gV)^N}{N!}\left(\frac{m}{2\pi}\right)^{\frac{3}{2}N}\frac{E^{\frac{3}{2}N-1}}{\Gamma\left(\frac{3}{2}N\right)}, \tag{122}$$

where m is the particle mass. Then the entropies of the ideal gas for the Tsallis and Boltzmann-Gibbs statistics are calculated by Eqs. (112), (113), and (122).

The temperatures (Eqs. (114) and (115)) for the ideal gas for the Tsallis and Boltzmann-Gibbs statistics in the microcanonical ensemble correspond to [6]

$$T = \frac{EW^{1/z}}{3N/2-1} \qquad \text{for } |z| < \infty, \tag{123}$$

$$T = \frac{E}{3N/2-1} \qquad \text{for } |z| = \infty. \tag{124}$$

The pressures (Eqs. (116) and (117)) and the chemical potentials (Eqs. (118) and (119)) for the ideal gas for the Tsallis and Boltzmann-Gibbs statistics in the microcanonical ensemble can be written as [6]

$$p = \frac{N}{V}\frac{E}{3N/2-1} \qquad \text{for } |z| < \infty \text{ and } |z| = \infty, \tag{125}$$

$$\mu = -\frac{E}{3N/2-1}\left[\ln\left(gV\left(\frac{mE}{2\pi}\right)^{3/2}\right) - \psi(N+1) - \frac{3}{2}\psi\left(\frac{3}{2}N\right)\right] \qquad \text{for } |z| < \infty \text{ and } |z| = \infty. \tag{126}$$

The variable (Eq. (120)) for the ideal gas for the Tsallis statistics in the microcanonical ensemble is [6]

$$\Xi = -\frac{E}{3N/2-1}\left[1 - W^{1/z} + \ln W^{1/z}\right] \qquad \text{for } |z| < \infty. \tag{127}$$

However, the variable (Eq. (121)) for the Boltzmann-Gibbs statistics vanishes, $\Xi = 0$.

4.2.2. Nonrelativistic ideal gas in the thermodynamic limit: microcanonical ensemble

Let us rewrite the thermodynamic quantities of the nonrelativistic ideal gas in the microcanonical ensemble in the terms of the thermodynamic limit when the entropic parameter z is considered to be an extensive variable of state

$$\begin{aligned} &E \to \infty,\ V \to \infty,\ N \to \infty,\ |z| \to \infty, \\ &\varepsilon = \frac{E}{N} = const,\ v = \frac{V}{N} = const,\ \tilde{z} = \frac{z}{N} = const. \end{aligned} \tag{128}$$

Then in the thermodynamic limit (Eq. (128)), the statistical weight (Eq. (122)) for the nonrelativistic ideal gas can be rewritten as [6]

$$W = w^N, \qquad w \equiv gv\left(\frac{m\varepsilon e^{5/3}}{3\pi}\right)^{3/2}. \tag{129}$$

Substituting Eq. (129) into Eqs. (112) and (113) and using Eq. (128), we obtain the entropy as [6]

$$S = N\tilde{z}\left(1 - w^{-1/\tilde{z}}\right) \qquad \text{for } |\tilde{z}| < \infty, \tag{130}$$

$$S = N \ln w \qquad \text{for } |\tilde{z}| = \infty. \tag{131}$$

The temperatures (Eqs. (123) and (124)) for the nonrelativistic ideal gas in the thermodynamic limit (128) can be rewritten as [6]

$$T = \frac{2}{3}\varepsilon w^{1/\tilde{z}} \qquad \text{for } |\tilde{z}| < \infty, \tag{132}$$

$$T = \frac{2}{3}\varepsilon \qquad \text{for } |\tilde{z}| = \infty. \tag{133}$$

The pressure (125) and the chemical potential (126) for the nonrelativistic ideal gas in the thermodynamic limit (128) become [6]

$$p = \frac{2}{3}\frac{\varepsilon}{v} \qquad \text{for } |\tilde{z}| < \infty \text{ and } |\tilde{z}| = \infty, \tag{134}$$

$$\mu = \frac{2}{3}\varepsilon\left(\frac{5}{2} - \ln w\right) \qquad \text{for } |\tilde{z}| < \infty \text{ and } |\tilde{z}| = \infty. \tag{135}$$

The variable (Eq. (127)) for the Tsallis statistics in the thermodynamic limit (Eq. (128)) corresponds to [6]

$$\Xi = -\frac{2}{3}\varepsilon\left[1 - w^{1/\tilde{z}} + \ln w^{1/\tilde{z}}\right] \qquad \text{for } |\tilde{z}| < \infty. \tag{136}$$

http://dx.doi.org/10.5772/60997

For the Boltzmann-Gibbs statistics, we have $\Xi=0$. Using the results so far obtained for the Tsallis statistics given in Eqs. (130), (132), and (134)-(136), we can verify that the Euler theorem defined in Eqs. (22) and (98) is satisfied [6], i.e., $TS=E+pV-\mu N+\Xi z$.

Let us verify the principle of additivity for the nonrelativistic ideal gas in the microcanonical ensemble in the thermodynamic limit when the entropic parameter z is an extensive variable of state. Suppose that the system is divided into two subsystems (1 and 2). Then the extensive variables of state of the microcanonical ensemble are additive [6]

$$E^{1+2}=E^1+E^2,\quad V^{1+2}=V^1+V^2,\quad N^{1+2}=N^1+N^2,\quad z^{1+2}=z^1+z^2. \tag{137}$$

However, the specific variables of state (Eq. (128)) are the same in each subsystem (intensive)

$$\varepsilon^{1+2}=\varepsilon^1=\varepsilon^2,\quad v^{1+2}=v^1=v^2,\quad \tilde{z}^{1+2}=\tilde{z}^1=\tilde{z}^2. \tag{138}$$

Considering Eqs. (129), (137), and (138), we can verify that the Tsallis thermodynamic potential (Eq. (130)) of the microcanonical ensemble is a homogeneous function of the first order, i.e., $S(E,\,V,\,N,\,z)/N=S(\varepsilon,\,v,\,\tilde{z})$, and it is additive (extensive) [6]

$$S^{1+2}\left(E^{1+2},V^{1+2},N^{1+2},z^{1+2}\right)=S^1\left(E^1,V^1,N^1,z^1\right)+S^2\left(E^2,V^2,N^2,z^2\right). \tag{139}$$

Now, considering Eqs. (129), (137), and (138), we find that the Tsallis temperature (Eq. (132)), the pressure (Eq. (134)), the chemical potential (Eq. (135)), and the variable (Eq. (136)) are the homogeneous functions of the zero order, i.e., $T(E,\,V,\,N,\,z)=T(\varepsilon,\,v,\,\tilde{z})$, $p(E,\,V,\,N,\,z)=p(\varepsilon,\,v,\,\tilde{z})$, $\mu(E,\,V,\,N,\,z)=\mu(\varepsilon,\,v,\,\tilde{z})$, and $\Xi(E,\,V,\,N,\,z)=\Xi(\varepsilon,\,v,\,\tilde{z})$, respectively, and they are the same in each subsystem (intensive) [6]

$$T^{1+2}\left(E^{1+2},V^{1+2},N^{1+2},z^{1+2}\right)=T^1\left(E^1,V^1,N^1,z^1\right)=T^2\left(E^2,V^2,N^2,z^2\right), \tag{140}$$

$$p^{1+2}\left(E^{1+2},V^{1+2},N^{1+2},z^{1+2}\right)=p^1\left(E^1,V^1,N^1,z^1\right)=p^2\left(E^2,V^2,N^2,z^2\right), \tag{141}$$

$$\mu^{1+2}\left(E^{1+2},V^{1+2},N^{1+2},z^{1+2}\right)=\mu^1\left(E^1,V^1,N^1,z^1\right)=\mu^2\left(E^2,V^2,N^2,z^2\right), \tag{142}$$

$$\Xi^{1+2}\left(E^{1+2},V^{1+2},N^{1+2},z^{1+2}\right)=\Xi^1\left(E^1,V^1,N^1,z^1\right)=\Xi^2\left(E^2,V^2,N^2,z^2\right). \tag{143}$$

Thus, the principle of additivity (Eqs. (21), (24), and (25)) is totally satisfied by the Tsallis statistics in the microcanonical ensemble. Equation (140) proves the zero law of thermodynamics for the microcanonical ensemble [6].

4.3. Equivalence of canonical and microcanonical ensembles

We can now easily prove the equivalence of the canonical and microcanonical ensembles for the Tsallis statistics in the thermodynamic limits (Eqs. (82) and (128)). Using Eqs. (83) and (129), it is easy to verify that Eq. (132) for the temperature of the microcanonical ensemble and Eq. (86) for the energy of canonical ensemble are identical. Comparing Eqs. (83) and (129) and using Eq. (86), we have

$$w = \left(\tilde{Z}_G e^{3/2}\right)^{\frac{\tilde{z}}{\tilde{z}+\frac{3}{2}}}. \tag{144}$$

Substituting Eq. (144) into Eq. (130) for the entropy of the microcanonical ensemble, we obtain the entropy of the canonical ensemble (Eq. (90)). Equation (134) for the pressure of the microcanonical ensemble is identical to Eq. (92) for the pressure of the canonical ensemble. Substituting Eqs. (144) and (86) into Eq. (135) for the chemical potential of the microcanonical ensemble, we obtain Eq. (94) for the chemical potential of the canonical ensemble. Moreover, substituting Eqs. (144) and (86) into Eq. (136) for the variable Ξ of the microcanonical ensemble, we obtain Eq. (96) for the variable Ξ of the canonical ensemble. Thus, for the Tsallis statistics, the canonical and microcanonical ensembles are equivalent in the thermodynamic limit when the entropic parameter z is considered to be an extensive variable of state.

5. Conclusions

In conclusion, let us summarize the main principles of the equilibrium statistical mechanics based on the generalized statistical entropy. The basic idea is that in the thermodynamic equilibrium, there exists a universal function called thermodynamic potential that completely describes the properties and states of the thermodynamic system. The fundamental thermodynamic potential, its arguments (variables of state), and its first partial derivatives with respect to the variables of state determine the complete set of physical quantities characterizing the properties of the thermodynamic system. The physical system can be prepared in many ways given by the different sets of the variables of state and their appropriate thermodynamic potentials. The first thermodynamic potential is obtained from the fundamental thermodynamic potential by the Legendre transform. The second thermodynamic potential is obtained by the substitution of one variable of state with the fundamental thermodynamic potential. Then the complete set of physical quantities and the appropriate thermodynamic potential determine the physical properties of the given system and their dependences. In the equilibrium thermodynamics, the thermodynamic potential of the physical system is given a priori, and it is a multivariate function of several variables of state. However, in the equilibrium

statistical mechanics, the thermodynamic potential is a composed function that can depend on the set of independent variables of state explicitly and implicitly through the probabilities of microstates. The probabilities of microstates are determined from the second part of the second law of thermodynamics, i.e., the maximum entropy principle. The equilibrium probability distributions are found from the constrained extremum of the thermodynamic potential as a function of a multidimensional set of probabilities considering that the statistical entropy is defined. The equilibrium thermodynamics and statistical mechanics are defined only on the class of homogeneous functions, i.e., all thermodynamic quantities describing the thermodynamic system should belong to the class of homogeneous functions of the first or zero orders.

In the present work, the general mathematical scheme of construction of the equilibrium statistical mechanics on the basis of an arbitrary definition of statistical entropy for two types of thermodynamic potential, the first and the second thermodynamic potentials, was proposed. As an example, we investigated the Tsallis and Boltzmann-Gibbs statistical entropies in the canonical and microcanonical ensembles. On the example of a nonrelativistic ideal gas, it was proven that the statistical mechanics based on the Tsallis entropy satisfies the requirements of the equilibrium thermodynamics only in the thermodynamic limit when the entropic index z is an extensive variable of state of the system. In this case the thermodynamic quantities of the Tsallis statistics belong to one of the classes of homogeneous functions of the first or zero orders.

Acknowledgements

This work was supported in part by the joint research project of JINR and IFIN-HH, protocol N 4342.

Author details

A.S. Parvan[1,2,3]

Address all correspondence to: parvan@theor.jinr.ru

1 Bogoliubov Laboratory of Theoretical Physics, Joint Institute for Nuclear Research, Dubna, Russian Federation

2 Department of Theoretical Physics, Horia Hulubei National Institute of Physics and Nuclear Engineering, Bucharest-Magurele, Romania

3 Institute of Applied Physics, Moldova Academy of Sciences, Chisinau, Republic of Moldova

References

[1] Gibbs J W. Elementary Principles in Statistical Mechanics, Developed with Especial Reference to the Rational Foundation of Thermodynamics. New Haven: Yale University Press; 1902

[2] Balescu R. Equilibrium and Nonequilibrium Statistical Mechanics. New York: Wiley; 1975.

[3] Tsallis C. Possible generalization of Boltzmann-Gibbs statistics. J. Stat. Phys. 1988;52(1-2):479-487.

[4] Tsallis C, Mendes R S, Plastino A R. The role of constraints within generalized nonextensive statistics. Physica A. 1998;261(3-4):534-554.

[5] Abe S, Martinez S, Pennini F, Plastino A. Classical gas in nonextensive optimal Lagrange multipliers formalism. Phys. Lett. A. 2001;278(5):249-254.

[6] Parvan A S. Microcanonical ensemble extensive thermodynamics of Tsallis statistics. Phys. Lett. A. 2006;350(5-6):331-338.

[7] Parvan A S. Extensive statistical mechanics based on nonadditive entropy: Canonical ensemble. Phys. Lett. A. 2006;360(1):26-34.

[8] Lenzi E K, Mendes R S, da Silva L R. Statistical mechanics based on Renyi entropy. Physica A. 2000;280(3-4):337-345.

[9] Parvan A S, Biro T S. Extensive Rényi statistics from non-extensive entropy. Phys. Lett. A. 2005;340(5-6):375-387.

[10] Parvan A S, Biro T S. Rényi statistics in equilibrium statistical mechanics. Phys. Lett. A. 2010;374(19-20):1951-1957.

[11] Wang Q A. Incomplete statistics: nonextensive generalizations of statistical mechanics. Chaos, Solitons and Fractals. 2001;12(8):1431-1437.

[12] Parvan A S, Biro T S. Equilibrium statistical mechanics for incomplete nonextensive statistics. Phys. Lett. A. 2011;375(3):372-378.

[13] Jaynes E T. Information Theory and Statistical Mechanics. II. Phys. Rev. 1957;108(2): 171.

[14] Prigogine I, Kondepudi D. Modern Thermodynamics: From Heat Engines to Dissipative Structures. Chichester: John Wiley & Sons; 1998.

[15] Kvasnikov I A. Thermodynamics and Statistical Mechanics: The Equilibrium Theory. Moscow: Moscow State Univ. Publ.; 1991.

[16] Botet R, Ploszajczak M, Gonzalez J A. Phase transitions in nonextensive spin systems. Phys. Rev. E. 2001;65(1):015103(R).

http://dx.doi.org/10.5772/60997

[17] Botet R, Ploszajczak M, Gudima K K, Parvan A S, Toneev V D. The thermodynamic limit in the non-extensive thermostatistics. Physica A. 2004;344(3-4):403-408.

[18] Krasnov M L, Makarenko G I, Kiseliov A I. Calculus of Variations: Problems and Exercises with detailed solutions. 2nd ed. Moscow: URSS Publisher; 2002.

[19] Arnold V I. Mathematical methods of classical mechanics. 2nd ed. New York : Springer-Verlag; 1989.

[20] Prigogine I, Defay R. Chemical Thermodynamics. London: Longmans, Green & Co Ltd; 1954.

Edited by Mofid Gorji-Bandpy

Thermodynamics is a branch of physics concerned with heat and temperature and their relation to energy and work. It defines macroscopic variables, such as internal energy, entropy, and pressure, that partly describe a body of matter or radiation. It states that the behavior of these variables is subject to general constraints that are common to all materials, not to the peculiar properties of particular materials. These general constraints are expressed in the three laws of thermodynamics which had a deep influence on the development of physics and chemistry. The book aims to present novel ideas that are crossing traditional disciplinary boundaries and introducing a wide spectrum of viewpoints and approaches in applied thermodynamics of the third millennium. The book will be of interest to those working in the fields of propulsion systems, power generation systems, chemical industry, quantum systems, refrigeration, fluid flow, combustion, and other phenomena.

ISBN 978-953-51-6650-4

Photo by weltreisendertj / DollarPhoto

IntechOpen