Advances in Modern Cosmology

Edited by Adnan Ghribi

ADVANCES IN MODERN COSMOLOGY

Edited by **Adnan Ghribi**

Advances in Modern Cosmology
http://dx.doi.org/10.5772/919
Edited by Adnan Ghribi

Contributors

Joel Berrier, Marc S. Seigar, Jean-Luc Starck, Marc Lachièze-Rey, Antoine Labatie, Biesiada, Sergei V. Ketov, Daniel Carter, Csaba Balazs, Michael Ibison, Michael Disney, Fabio L. Braghin, Bikash Chandra Paul

Notice

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

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

4,000+
Open access books available

116,000+
International authors and editors

120M+
Downloads

151
Countries delivered to

Our authors are among the
Top 1%
most cited scientists

12.2%
Contributors from top 500 universities

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Meet the editor

Dr. Adnan Ghribi got his Master degree in Space Engineering from University Paul Sabatier in Toulouse France in 2006 and his PhD in Observational Cosmology from University Denis Diderot in Paris France in 2009. His main research topic is the study of the polarization of the Cosmic Microwave Background and the detection of the primordial gravitational waves. He worked in the AstroParticule and Cosmology laboratory in Paris on the Q and U Bolometric Interferometer for Cosmology. He continues his research in the University of California Berkeley on the PolarBear instrument and the projects PolarBear II and LiteBird.

Contents

Preface

To be human is to care how the physical world came, whether it has boundaries and what is to become of it. Cosmology is the science that tries to answer these eternal questions. During the twentieth century, it has been elevated from the rank of philosophy to precision science thanks to the advances in both theory and observation. General relativity, quantum mechanics and observational techniques gave birth to the modern cosmology. The family of models that are known to explain the best the observations is the Cold Dark Matter model with dark energy also known as the standard model or ΛCDM. The ΛCDM model opened the door for several cosmology subfields like the study of the very early Universe, Big-Bang nucleosynthesis, Cosmic Microwave Background (CMB), formation and evolution of large scale structures, dark matter and dark energy. According to the observation of galaxies and CMB (relic radiation emitted in the early ages of the Universe), dark matter accounts for 23% of the mass energy density of the observable Universe while ordinary matter accounts only for 4.6%. The remainder is attributed to dark energy. That is that, today, nor do we know what constitutes 83% of the matter in the Universe (dark matter), neither do we understand the nature of the energy that accelerates the expansion of the Universe (dark energy). This book focus on the these unanswered question while providing an overview of some of the most promising advances in modern cosmology.

In its first part, the book focus on dark matter. Extensions of the standard model are proposed by introducing the supersymmetric dark matter and local supersymmetry, also known as supergravity. Other investigations of large scale and galaxy scale structures attempt to explain and understand the nature and distribution of dark matter. The second part of the book is about the problem of dark energy. Several models try to understand the nature of dark energy. One of them, the holographic dark energy model with modified variable Chaplygin gas, is detailed in chapter 5. In chapter 6, strong lensing systems are considered as possible observational probes for dark energy models. The seventh chapter is a theoretical investigation of the effect of the expansion of the Universe in the context of general relativity on electromagnetic radiation and fermionic matter. Finally, the last chapter is a review of the different

methods that allow to extract the information from galaxy surveys for the study of large scale structures.

Dr. Adnan Ghribi
Experimental Cosmology Group
University of California Berkeley
USA

Part 1

Dark Matter

$F(\mathcal{R})$ **Supergravity and Early Universe: the Meeting Point of Cosmology and High-Energy Physics**

Sergei V. Ketov

Department of Physics, Tokyo Metropolitan University, Minami-ohsawa 1-1,
Hachioji-shi, Tokyo 192-0397
Institute for the Physics and Mathematics of the Universe (IPMU), The University of
Tokyo, Kashiwanoha 5-1-5, Kashiwa-shi, Chiba 277-8568
Japan

1. Introduction

In this Chapter we focus on the field-theoretical description of the inflationary phase of the early universe and its post-inflationary dynamics (reheating and particle production) in the context of supergravity, based on the original papers (1–10). To begin with, let us first introduce some basics of inflation.

Cosmological inflation (a phase of 'rapid' quasi-exponential accelerated expansion of universe) (11–13) predicts homogeneity of our Universe at large scales, its spatial flatness, large size and entropy, and the almost scale-invariant spectrum of cosmological perturbations, in good agreement with the WMAP measurements of the CMB radiation spectrum (14; 15). Inflation is also the only known way to generate structure formation in the universe via amplifying quantum fluctuations in vacuum.

However, inflation is just the cosomological paradigm, not a theory! The known field-theoretical mechanisms of inflation use a slow-roll scalar field ϕ (called inflaton) with proper scalar potential $V(\phi)$ (12; 13).

The scale of inflation is well beyond the electro-weak scale, ie. is well beyond the Standard Model of Elementary Particles! Thus the inflationary stage in the early universe is the most powerful High-Energy Physics (HEP) accelerator in Nature (up to 10^{10} *TeV*). Therefore, inflation is the great and unique window to HEP!

The nature of inflaton and the origin of its scalar potential are the big mysteries.

Throughout the paper the units $\hbar = c = 1$ and the spacetime signature $(+, -, -, -)$ are used. See ref. (16) for our use of Riemann geometry of a curved spacetime.

The Cosmic Microwave Background (CMB) radiation from the Wilkinson Microwave Anisotropy Probe (WMAP) satellite mission (14) is one of the main sources of data about the early universe. Deciphering the CMB in terms of the density perturbations, gravity wave polarization, power spectrum and its various indices is a formidable task. It also requires the heavy CMB mathematical formalism based on General Relativity — see eg., the textbooks (17–19). Fortunately, we do not need that formalism for our purposes, since the relevant indices can also be introduced in terms of the inflaton scalar potential (Sec. 4). We assume

that inflation did happen. There exist many inflationary models — see eg. the textbook (13) for their description and comparison (without supersymmetry). Our aim is a viable theoretical description of inflation in the context of supergravity.

The main Cosmological Principle of a spatially homogeneous and isotropic $(1 + 3)$-dimensional universe (at large scales) gives rise to the FLRW metric

$$ds^2_{\text{FLRW}} = dt^2 - a^2(t) \left[\frac{dr^2}{1 - kr^2} + r^2 d\Omega^2 \right] \tag{1}$$

where the function $a(t)$ is known as the scale factor in 'cosmic' (comoving) coordinates (t, r, θ, ϕ), and k is the FLRW topology index, $k = (-1, 0, +1)$. The FLRW metric (1) admits the six-dimensional isometry group G that is either $SO(1, 3)$, $E(3)$ or $SO(4)$, acting on the orbits $G/SO(3)$, with the spatial three-dimensional sections H^3, E^3 or S^3, respectively. The Weyl tensor of any FLRW metric vanishes,

$$C^{\text{FLRW}}_{\mu\nu\lambda\rho} = 0 \tag{2}$$

where $\mu, \nu, \lambda, \rho = 0, 1, 2, 3$. The early universe inflation (acceleration) means

$$\ddot{a}(t) > 0 \text{ , or equivalently , } \frac{d}{dt} \left(\frac{H^{-1}}{a} \right) < 0 \tag{3}$$

where $H = \dot{a}/a$ is called Hubble function. We take $k = 0$ for simplicity. The amount of inflation (called the *e-foldings* number) is given by

$$N_e = \ln \frac{a(t_{\text{end}})}{a(t_{\text{start}})} = \int_{t_{\text{start}}}^{t_{\text{end}}} H \, dt \approx \frac{1}{M_{\text{Pl}}^2} \int_{\phi_{\text{end}}}^{\phi} \frac{V}{V'} \, d\phi \tag{4}$$

Next, a few words about our strategy. It is well recognized now that one has to go beyond the Einstein-Hilbert action for gravity, both from the experimental viewpoint (eg.,because of Dark Energy) and from the theoretical viewpoint (eg., because of the UV incompleteness of quantized Einstein gravity, and the need of its unification with the Standard Model of Elementary Particles).

In our approach, the origin of inflation is purely *geometrical*, ie. is closely related to space-time and gravity. It can be technically accomplished by taking into account the higher-order curvature terms on the left-hand-side of Einstein equations, and extending gravity to supergravity. The higher-order curvature terms are supposed to appear in the gravitational effective action of Quantum Gravity. Their derivation from Superstring Theory may be possible too. The true problem is a *selection* of those high-order curvature terms that are physically relevant or derived from a fundamental theory of Quantum Gravity.

There are many phenomenological models of inflation in the literature, which usually employ some new fields and new interactions. It is, therefore, quite reasonable and meaningful to search for the *minimal* inflationary model building, by getting most economical and viable inflationary scenario. I am going to use the one proposed the long time ago by Starobinsky (20; 21), which does not use new fields (beyond a spacetime metric) and exploits only gravitational interactions. I also assume that the general coordinate invariance in spacetime is fundamental, and it should not be sacrificed. Moreover, it should be extended to the more fundamental, local supersymmetry that is known to imply the general coordinate invariance.

On the theoretical side, the available inflationary models may be also evaluated with respect to their *"cost"*, ie. against what one gets from a given model in relation to what one puts in! Our approach does *not* introduce new fields, beyond those already present in gravity and supergravity. We also exploit (super)gravity interactions *only*, ie. do *not* introduce new interactions, in order to describe inflation.

Before going into details, let me address two common prejudices and objections.

The higher-order curvature terms are usually expected to be relevant near the spacetime curvature singularities. It is also quite possible that some higher-derivative gravity, subject to suitable constraints, could be the effective action to a quantized theory of gravity, [1] like eg., in String Theory. However, there are also some common doubts against the higher-derivative terms, in principle.

First, it is often argued that all higher-derivative field theories, including the higher-derivative gravity theories, have ghosts (i.e. are unphysical), because of Ostrogradski theorem (1850) in Classical Mechanics. As a matter of fact, though the presence of ghosts is a generic feature of the higher-derivative theories indeed, it is not always the case, while many explicit examples are known (Lovelock gravity, Euler densities, some $f(R)$ gravity theories, etc.) — see eg., ref. (22) for more details. In our approach, the absence of ghosts and tachyons is required, and is considered as one of the main physical selection criteria for the good higher-derivative field theories.

Another common objection against the higher-derivative gravity theories is due to the fact that all the higher-order curvature terms in the action are to be suppressed by the inverse powers of M_{Pl} on dimensional reasons and, therefore, they seem to be 'very small and negligible'. Though it is generically true, it does not mean that all the higher-order curvature terms are irrelevant at all scales much less than M_{Pl}. For instance, it appears that the *quadratic* curvature terms have *dimensionless* couplings, while they can be instrumental for an early universe inflation. A non-trivial function of R in the effective gravitational action may also 'explain' the Dark Energy phenomenon in the present Universe.

Cosmological inflation in supergravity is a window to High-Energy Physics beyond the Standard Model of Elementary Particles. The Starobinsky inflationary model is introduced in Sec. 2. Its classical equivalence to a scalar-tensor gravity is shown in Sec. 3, and its observational predictions for the CMB are given in Sec. 4. We review a construction of the new $F(\mathcal{R})$ supergravity theories in Secs. 5 and 6. The $F(\mathcal{R})$ supergravity theories are the $N = 1$ locally supersymmetric extensions of the well studied $f(R)$ gravity theories in four space-time dimensions, which are often used for 'explaining' inflation and Dark Energy. A manifeslty supersymmetric description of the $F(\mathcal{R})$ supergravities exist in terms of $N = 1$ superfields, by using the (old) minimal Poincaré supergravity in curved superspace. We prove that any $F(\mathcal{R})$ supergravity is classically equivalent to the particular Poincaré-type matter-coupled $N = 1$ supergravity via the superfield Legendre-Weyl-Kähler transformation. The (nontrivial) Kähler potential and the scalar superpotential of inflaton superfield are determined in terms of the original holomorphic $F(\mathcal{R})$ function. The conditions for stability, the absence of ghosts and tachyons are also found. No-scale $F(\mathcal{R})$ supergravity is constructed too (Sec. 7). Three different examples of the $F(\mathcal{R})$ supergravity theories are studied in detail. The first example is devoted to recovery of the standard (pure) $N = 1$ supergravity with a negative cosmological constant from $F(\mathcal{R})$ supergravity (Sec. 8). As the second example, a generic $\mathcal{R}^2$ supergravity is investigated, the existence of the AdS bound on the scalar curvature and a possibility of positive cosmological constant are discovered (Sec. 9). As

---

[1] To the best of my knowledge, this proposal was first formulated by A.D. Sakharov in 1967.

the third example, a simple and viable realization of chaotic inflation in supergravity is given, via an embedding of the Starobinsky inflationary model into the $F(\mathcal{R})$ supergravity (Sec. 10). Our approach does not introduce new exotic fields or new interactions, beyond those already present in (super)gravity. In Sec. 11 the nonminimal scalar-curvature couplings in gravity and supergravity, and their correspondence to $f(R)$ gravity and $F(\mathcal{R})$ supergravity, respectively, are analyzed within slow-roll inflation. Reheating and particle production are briefly discussed in Sec. 12. Our short conclusion is Sec. 13. In our outlook (Sec. 14), we emphasize the possible use of $F(\mathcal{R})$ supergravity towards solving the outstanding problems of CP-violation, the origin of baryonic asymmetry, lepto- and baryo-genesis.

2. Starobinsky minimal model of inflation

It can be argued that it is the *scalar* curvature-dependent part of the gravitational effective action that is most relevant to the large-scale dynamics $H(t)$. Here are some simple arguments. In 4 dimensions all the independent *quadratic* curvature invariants are $R^{\mu\nu\lambda\rho}R_{\mu\nu\lambda\rho}$, $R^{\mu\nu}R_{\mu\nu}$ and R^2. However,

$$\int d^4x \sqrt{-g} \left(R^{\mu\nu\lambda\rho}R_{\mu\nu\lambda\rho} - 4R^{\mu\nu}R_{\mu\nu} + R^2 \right) \tag{5}$$

is topological (ie. a total derivative) for any metric, while

$$\int d^4x \sqrt{-g} \left(3R^{\mu\nu}R_{\mu\nu} - R^2 \right) \tag{6}$$

is also topological for any FLRW metric, because of eq. (2). Hence, the FLRW-relevant quadratically-generated gravity action is $(8\pi G_N = 1)$

$$S = -\frac{1}{2} \int d^4x \sqrt{-g} \left(R - R^2/M^2 \right) \tag{7}$$

This action is known as the Starobinsky model (20; 21). Its equations of motion allow a stable inflationary solution, and it is an attractor! In particular, for $H \gg M$, one finds

$$H \approx \left(\frac{M}{6} \right)^2 (t_{\text{end}} - t) \tag{8}$$

It is the particular realization of chaotic inflation (ie. with chaotic initial conditions) (23), and with a Graceful Exit.

In the case of a generic gravitational action with the higher-order curvature terms, the Weyl dependence can be excluded due to eq. (2) again. A dependence upon the Ricci tensor can be also excluded since, otherwise, it would lead to the extra propagating massless spin-2 degree of freedom (in addition to a metric) described by the field $\partial\mathcal{L}/\partial R_{\mu\nu}$. The higher derivatives of the scalar curvature in the gravitational Lagrangian $\mathcal{L}$ just lead to more propagating scalars (24), so I simply ignore them for simplicity in what follows.

3. $f(R)$ Gravity and scalar-tensor gravity

The Starobinsky model (7) is the special case of the $f(R)$ gravity theories (25; 26) having the action

$$S_f = -\frac{1}{16\pi G_N} \int d^4x \sqrt{-g} \, \tilde{f}(R) \tag{9}$$

In the absence of extra matter, the gravitational (trace) equation of motion is of the fourth order with respect to the time derivative,

$$\frac{3}{a^3}\frac{d}{dt}\left(a^3 \frac{d\tilde{f}'(R)}{dt}\right) + R\tilde{f}'(R) - 2\tilde{f}(R) = 0 \tag{10}$$

where we have used $H = \frac{\dot{a}}{a}$ and $R = -6(\dot{H} + 2H^2)$. The primes denote the derivatives with respect to R, and the dots denote the derivative with respect to t. Static de-Sitter solutions correspond to the roots of the equation $R\tilde{f}'(R) = 2\tilde{f}(R)$ (27).

The 00-component of the gravitational equations is of the third order with respect to the time derivative,

$$3H\frac{d\tilde{f}'(R)}{dt} - 3(\dot{H} + H^2)\tilde{f}'(R) - \frac{1}{2}\tilde{f}(R) = 0 \tag{11}$$

The (classical and quantum) *stability* conditions in $f(R)$ gravity are well known (25; 26), and are given by (in our notation)

$$\tilde{f}'(R) > 0 \qquad \text{and} \qquad \tilde{f}''(R) < 0 \tag{12}$$

respectively. The first condition (12) is needed to get a physical (non-ghost) graviton, while the second condition (12) is needed to get a physical (non-tachyonic) scalaron (see Sec. 9 for more).

Any $f(R)$ gravity is known to be classically *equivalent* to the certain scalar-tensor gravity having an (extra) propagating scalar field (28–30). The formal equivalence can be established via a Legendre-Weyl transform.

First, the $f(R)$-gravity action (9) can be rewritten to the form

$$S_A = \frac{-1}{2\kappa^2} \int d^4x \sqrt{-g} \left\{AR - Z(A)\right\} \tag{13}$$

where the real scalar (or Lagrange multiplier) $A(x)$ is related to the scalar curvature R by the Legendre transformation:

$$R = Z'(A) \qquad \text{and} \qquad \tilde{f}(R) = RA(R) - Z(A(R)) \tag{14}$$

with $\kappa^2 = 8\pi G_N = M_{\text{Pl}}^{-2}$.

Next, a Weyl transformation of the metric,

$$g_{\mu\nu}(x) \rightarrow \exp\left[\frac{2\kappa\phi(x)}{\sqrt{6}}\right] g_{\mu\nu}(x) \tag{15}$$

with arbitrary field parameter $\phi(x)$ yields

$$\sqrt{-g}R \rightarrow \sqrt{-g}\exp\left[\frac{2\kappa\phi(x)}{\sqrt{6}}\right] \left\{R - \sqrt{\frac{6}{-g}}\partial_\mu\left(\sqrt{-g}g^{\mu\nu}\partial_\nu\phi\right)\kappa - \kappa^2 g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi\right\} \tag{16}$$

Therefore, when choosing

$$A(\kappa\phi) = \exp\left[\frac{-2\kappa\phi(x)}{\sqrt{6}}\right] \tag{17}$$

and ignoring a total derivative in the Lagrangian, we can rewrite the action to the form

$$S[g_{\mu\nu}, \phi] = \int d^4x \sqrt{-g} \left\{ \frac{-R}{2\kappa^2} + \frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi \right. \\ \left. + \frac{1}{2\kappa^2} \exp\left[\frac{4\kappa\phi(x)}{\sqrt{6}} \right] Z(A(\kappa\phi)) \right\} \tag{18}$$

in terms of the physical (and canonically normalized) scalar field $\phi(x)$, without any higher derivatives and ghosts. As a result, one arrives at the standard action of the real dynamical scalar field $\phi(x)$ *minimally* coupled to Einstein gravity and having the scalar potential

$$V(\phi) = -\frac{M_{\text{Pl}}^2}{2} \exp\left\{ \frac{4\phi}{M_{\text{Pl}}\sqrt{6}} \right\} Z\left(\exp\left[\frac{-2\phi}{M_{\text{Pl}}\sqrt{6}} \right] \right) \tag{19}$$

In the context of the inflationary theory, the *scalaron* (= scalar part of spacetime metric) ϕ can be identified with inflaton. This inflaton has clear origin, and may also be understood as the conformal mode of the metric over Minkowski or (A)dS vacuum.

In the Starobinsky case of $\tilde{f}(R) = R - R^2/M^2$, the inflaton scalar potential reads

$$V(y) = V_0 \left(e^{-y} - 1 \right)^2 \tag{20}$$

where we have introduced the notation

$$y = \sqrt{\frac{2}{3}} \frac{\phi}{M_{\text{Pl}}} \qquad \text{and} \qquad V_0 = \frac{1}{8} M_{\text{Pl}}^2 M^2 \tag{21}$$

It is worth noticing here the appearance of the inflaton vacuum energy V_0 driving inflation. The end of inflation (Graceful Exit) is also clear: the scalar potential (20) has a very flat (slow-roll) 'plateau', ending with a 'waterfall' towards the minimum (Fig. 1).

It is worth emphasizing that the inflaton (scalaron) scalar potential (20) is derived here by merely assuming the existence of the R^2 term in the gravitational action. The Newton (weak gravity) limit is not applicable to an early universe (including its inflationary stage), so that the dimensionless coefficient in front of the R^2 term does not have to be very small. It distinguishes the primordial 'dark energy' driving inflation in the early Universe from the 'Dark Energy' responsible for the present Universe acceleration.

4. Inflationary theory and observations

The *slow-roll* inflation parameters are defined by

$$\varepsilon(\phi) = \frac{1}{2} M_{\text{Pl}}^2 \left(\frac{V'}{V} \right)^2 \qquad \text{and} \qquad \eta(\phi) = M_{\text{Pl}}^2 \frac{V''}{V} \tag{22}$$

A necessary condition for the slow-roll approximation is the smallness of the inflation parameters

$$\varepsilon(\phi) \ll 1 \qquad \text{and} \qquad |\eta(\phi)| \ll 1 \tag{23}$$

The first condition implies $\ddot{a}(t) > 0$. The second one guarantees that inflation lasts long enough, via domination of the *friction* term in the inflaton equation of motion, $3H\dot{\phi} = -V'$.
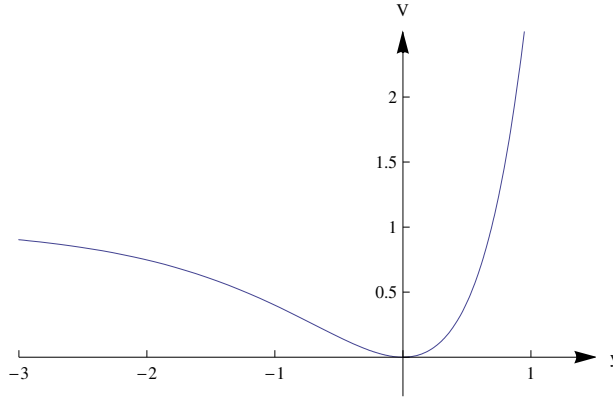
Fig. 1. The inflaton scalar potential $v(x) = (e^y - 1)^2$ in the Starobinsky model, after $y \to -y$

As is well known (13), scalar and tensor perturbations of the metric *decouple*. The scalar perturbations couple to the density of matter and radiation, so they are responsible for the inhomogeneities and anisotropies in the universe. The tensor perturbations (or gravity waves) also contribute to the CMB, while their experimental detection would tell us much more about inflation. The CMB radiation is expected to be *polarized* due to Compton scattering at the time of decoupling (31; 32).

The primordial spectrum is proportional to k^{n-1}, in terms of the comoving wave number k and the spectral index n. In theory, the slope n_s of the *scalar* power spectrum, associated with the density perturbations, $\left(\frac{\delta\rho}{\rho}\right)^2 \propto k^{n_s-1}$, is given by $n_s = 1 + 2\eta - 6\varepsilon$, the slope of the *tensor* primordial spectrum, associated with gravitational waves, is $n_t = -2\varepsilon$, and the *tensor-to-scalar ratio* is $r = 16\varepsilon$ (see eg., ref. (13)).

It is straightforward to calculate those indices in any inflationary model with a given inflaton scalar potential. In the case of the Starobinsky model and its scalar potential (20), one finds (6; 33; 34)

$$n_s = 1 - \frac{2}{N_e} + \frac{3\ln N_e}{2N_e^2} - \frac{2}{N_e^2} + \mathcal{O}\left(\frac{\ln^2 N_e}{N_e^3}\right) \tag{24}$$

and

$$r \approx \frac{12}{N_e^2} \approx 0.004 \tag{25}$$

with $N_e \approx 55$. The very small value of r is the sharp prediction of the Starobinsky inflationary model towards r-measurements in a future.

Those theoretical values are to be compared to the observed values of the CMB radiation due to the WMAP satellite mission. For instance, the most recent WMAP7 observations (14) yield

$$n_s = 0.963 \pm 0.012 \quad \text{and} \quad r < 0.24 \tag{26}$$

with the 95 % level of confidence.

The amplitude of the initial perturbations, $\Delta_R^2 = M_{\text{Pl}}^4 V / (24\pi^2\varepsilon)$, is also a physical observable, whose experimental value is known due to another Cosmic Background Explorer (COBE) satellite mission (35):

$$\left(\frac{V}{\varepsilon}\right)^{1/4} = 0.027 \, M_{\text{Pl}} = 6.6 \times 10^{16} \text{ GeV} \tag{27}$$
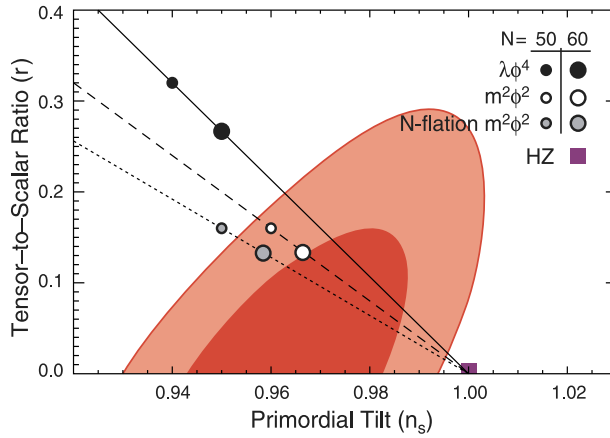
Fig. 2. Starobinsky inflation vs. $m^2\phi^2/2$ and $\lambda\phi^4$

It determines the normalization of the R^2-term in the action (7)

$$\frac{M}{M_{\text{Pl}}} = 4 \cdot \sqrt{\frac{2}{3}} \cdot (2.7)^2 \cdot \frac{e^{-y}}{(1-e^{-y})^2} \cdot 10^{-4} \approx (3.5 \pm 1.2) \cdot 10^{-6} \tag{28}$$

In particular, the inflaton mass is given by $M_{\text{inf}} = M/\sqrt{6}$.

The main theoretical lessons, that we can draw from the discussion above towards our next goals, are:

(i) the main *discriminants* amongst all inflationary models are given by the values of n_s and r;

(ii) the Starobinsky model (1980) of chaotic inflation is very simple and economic. It uses gravity interactions only. It predicts the origin of inflaton and its scalar potential. It is still *viable* and *consistent* with all known observations. Inflaton is not charged (singlet) under the SM gauge group. The Starobinsky inflation has an end (Graceful Exit), and gives the simple explanation to the WMAP-observed value of n_s. The *key* difference of Starobinsky inflation from the other standard inflationary models (having $\frac{1}{2}m^2\phi^2$ or $\lambda\phi^4$ scalar potentials) is the very low value of r — see the standard Fig. 2 for a comparison and ref. (36) for details. A discovery of primordial gravitational waves and precision measurements of the value of r (if $r \geq 0.1$) with the accuracy of 0.5% may happen due to the ongoing PLANCK satellite mission (37);

(iii) the viable inflationary models, based on $\tilde{f}(R) = R + \hat{f}(R)$ gravity, turn out to be close to the simplest Starobinsky model (over the range of R relevant to inflation), with $\hat{f}(R) \approx R^2 A(R)$ and the slowly varying function $A(R)$ in the sense

$$\left|A'(R)\right| \ll \frac{A(R)}{R} \qquad \text{and} \qquad \left|A''(R)\right| \ll \frac{A(R)}{R^2} \tag{29}$$

5. Supergravity and superspace

Supersymmetry (SUSY) is the symmetry between bosons and fermions. SUSY is the natural extension of Poincaré symmetry, and is well motivated in HEP beyond the SM. Supersymmetry is also needed for consistency of strings. Supergravity (SUGRA) is the theory of *local* supersymmetry that implies general coordinate invariance. In other words,

considering inflation with supersymmetry necessarily leads to supergravity. As a matter of fact, most of studies of superstring- and brane-cosmology are also based on their effective description in the 4-dimensional $N = 1$ supergravity.

It is not our purpose here to give a detailed account of SUSY and SUGRA, because of the existence of several textbooks — see e.g., refs. (38–40). In this Section I recall only the basic facts about $N = 1$ supergravity in four spacetime dimensions, which are needed here.

A concise and manifestly supersymmetric description of SUGRA is given by Superspace. In this section the natural units $c = \hbar = \kappa = 1$ are used.

Supergravity needs a curved superspace. However, they are not the same, because one has to reduce the field content to the minimal one corresponding to off-shell supergravity multiplets. It is done by imposing certain constraints on the supertorsion tensor in curved superspace (38–40). An off-shell supergravity multiplet has some extra (auxiliary) fields with noncanonical dimensions, in addition to physical spin-2 field (metric) and spin-3/2 field (gravitino). It is worth mentioning that imposing the off-shell constraints is independent upon writing a supergravity action.

One may work either in a full superspace or in a chiral one. There are certain anvantages of using the chiral superspace, because it helps us to keep the auxiliary fields unphysical (i.e. nonpropagating).

The chiral superspace density (in the supersymmetric gauge-fixed form) reads

$$\mathcal{E}(x,\theta) = e(x) \left[1 - 2i\theta\sigma_a \bar{\psi}^a(x) + \theta^2 B(x) \right] , \tag{30}$$

where $e = \sqrt{-\det g_{\mu\nu}}$, $g_{\mu\nu}$ is a spacetime metric, $\psi_\alpha^a = e_\mu^a \psi_\alpha^\mu$ is a chiral gravitino, $B = S - iP$ is the complex scalar auxiliary field. We use the lower case middle greek letters $\mu, \nu, \ldots = 0, 1, 2, 3$ for curved spacetime vector indices, the lower case early latin letters $a, b, \ldots = 0, 1, 2, 3$ for flat (target) space vector indices, and the lower case early greek letters $\alpha, \beta, \ldots = 1, 2$ for chiral spinor indices.

A solution to the superspace Bianchi identities together with the constraints defining the $N = 1$ Poincaré-type *minimal* supergravity theory results in only *three* covariant tensor superfields $\mathcal{R}$, $\mathcal{G}_a$ and $\mathcal{W}_{\alpha\beta\gamma}$, subject to the off-shell relations (38–40):

$$\mathcal{G}_a = \bar{\mathcal{G}}_a , \qquad \mathcal{W}_{\alpha\beta\gamma} = \mathcal{W}_{(\alpha\beta\gamma)} , \qquad \bar{\nabla}_{\dot{\alpha}} \mathcal{R} = \bar{\nabla}_{\dot{\alpha}} \mathcal{W}_{\alpha\beta\gamma} = 0 , \tag{31}$$

and

$$\bar{\nabla}^{\dot{\alpha}} \mathcal{G}_{\alpha\dot{\alpha}} = \nabla_\alpha \mathcal{R} , \qquad \nabla^\gamma \mathcal{W}_{\alpha\beta\gamma} = \tfrac{i}{2} \nabla_\alpha^{\dot{\alpha}} \mathcal{G}_{\beta\dot{\alpha}} + \tfrac{i}{2} \nabla_\beta^{\dot{\alpha}} \mathcal{G}_{\alpha\dot{\alpha}} , \tag{32}$$

where $(\nabla_\alpha, \bar{\nabla}_{\dot{\alpha}}, \nabla_{\alpha\dot{\alpha}})$ stand for the curved superspace $N = 1$ supercovariant derivatives, and the bars denote complex conjugation.

The covariantly chiral complex scalar superfield $\mathcal{R}$ has the scalar curvature R as the coefficient at its θ^2 term, the real vector superfield $\mathcal{G}_{\alpha\dot{\alpha}}$ has the traceless Ricci tensor, $R_{\mu\nu} + R_{\nu\mu} - \tfrac{1}{2} g_{\mu\nu} R$, as the coefficient at its $\theta\sigma^a\bar{\theta}$ term, whereas the covariantly chiral, complex, totally symmetric, fermionic superfield $\mathcal{W}_{\alpha\beta\gamma}$ has the self-dual part of the Weyl tensor $C_{\alpha\beta\gamma\delta}$ as the coefficient at its linear θ^δ-dependent term.

A generic Lagrangian representing the supergravitational effective action in (full) superspace, reads

$$\mathcal{L} = \mathcal{L}(\mathcal{R}, \mathcal{G}, \mathcal{W}, \ldots) \tag{33}$$

where the dots stand for arbitrary supercovariant derivatives of the superfields.

The Lagrangian (33) it its most general form is, however, unsuitable for physical applications, not only because it is too complicated, but just because it generically leads to propagating auxiliary fields, which break the balance of the bosonic and fermionic degrees of freedom. The important physical condition of keeping the supergravity auxiliary fields to be truly auxiliary (ie. nonphysical or nonpropagating) in field theories with the higher derivatives was dubbed the *'auxiliary freedom'* in refs. (41; 42). To get the supergravity actions with the 'auxiliary freedom', we will use a chiral (curved) superspace.

6. $F(\mathcal{R})$ supergravity in superspace

Let us first concentrate on the scalar-curvature-sector of a generic higher-derivative supergravity (33), which is most relevant to the FRLW cosmology, by ignoring the tensor curvature superfields $\mathcal{W}_{\alpha\beta\gamma}$ and $\mathcal{G}_{\alpha\dot{\alpha}}$, as well as the derivatives of the scalar superfield $\mathcal{R}$, like that in Sec. 2. Then we arrive at the chiral superspace action

$$S_F = \int d^4x d^2\theta \, \mathcal{E} F(\mathcal{R}) + \text{H.c.} \tag{34}$$

governed by a chiral or *holomorphic* function $F(\mathcal{R})$. [2] Besides having the manifest local $N = 1$ supersymmetry, the action (34) has the auxiliary freedom since the auxiliary field B does not propagate. It distinguishes the action (34) from other possible truncations of eq. (33). The action (34) gives rise to the spacetime *torsion* generated by gravitino, while its bosonic terms have the form

$$S_f = -\frac{1}{2} \int d^4x \, \sqrt{-g} \, \tilde{f}(R) \tag{35}$$

Hence, eq. (34) can also be considered as the locally $N = 1$ supersymmetric extension of the $f(R)$-type gravity (Sec. 3). However, in the context of supergravity, the choice of possible bosonic functions $\tilde{f}(R)$ is very restrictive (see Secs. 9 and 10).

The superfield action (34) is classically equivalent to

$$S_V = \int d^4x d^2\theta \, \mathcal{E} \left[\mathcal{Z}\mathcal{R} - V(\mathcal{Z}) \right] + \text{H.c.} \tag{36}$$

with the covariantly chiral superfield $\mathcal{Z}$ as the Lagrange multiplier superfield. Varying the action (36) with respect to $\mathcal{Z}$ gives back the original action (34) provided that

$$F(\mathcal{R}) = \mathcal{R}\mathcal{Z}(\mathcal{R}) - V(\mathcal{Z}(\mathcal{R})) \tag{37}$$

where the function $\mathcal{Z}(\mathcal{R})$ is defined by inverting the function

$$\mathcal{R} = V'(\mathcal{Z}) \tag{38}$$

Equations (37) and (38) define the superfield Legendre transform, and imply

$$F'(\mathcal{R}) = Z(\mathcal{R}) \quad \text{and} \quad F''(\mathcal{R}) = Z'(\mathcal{R}) = \frac{1}{V''(\mathcal{Z}(\mathcal{R}))} \tag{39}$$

where $V'' = d^2V/d\mathcal{Z}^2$. The second formula (39) is the *duality* relation between the supergravitational function F and the chiral superpotential V.

---

[2] The similar component field construction, by the use of the 4D, $N = 1$ superconformal tensor calculus, was given in ref. (43).

A supersymmetric (local) Weyl transform of the acton (36) can be done entirely in superspace. In terms of the field components, the super-Weyl transform amounts to a Weyl transform, a chiral rotation and a (superconformal) *S*-supersymmetry transformation (44). The chiral density superfield $\mathcal{E}$ appears to be the chiral *compensator* of the super-Weyl transformations,

$$\mathcal{E} \to e^{3\Phi} \mathcal{E} \tag{40}$$

whose parameter Φ is an arbitrary covariantly chiral superfield, $\bar{\nabla}_{\dot{\alpha}} \Phi = 0$. Under the transformation (40) the covariantly chiral superfield $\mathcal{R}$ transforms as

$$\mathcal{R} \to e^{-2\Phi} \left(\mathcal{R} - \tfrac{1}{4} \bar{\nabla}^2 \right) e^{\bar{\Phi}} \tag{41}$$

The super-Weyl chiral superfield parameter Φ can be traded for the chiral Lagrange multiplier $\mathcal{Z}$ by using a generic gauge condition

$$\mathcal{Z} = \mathcal{Z}(\Phi) \tag{42}$$

where $\mathcal{Z}(\Phi)$ is a holomorphic function of Φ. It results in the action

$$S_\Phi = \int d^4x d^4\theta \, E^{-1} e^{\Phi+\bar{\Phi}} \left[\mathcal{Z}(\Phi) + \text{H.c.} \right] - \int d^4x d^2\theta \, \mathcal{E} e^{3\Phi} V(\mathcal{Z}(\Phi)) + \text{H.c.} \tag{43}$$

Equation (43) has the *standard* form of the action of a chiral matter superfield coupled to supergravity,

$$S[\Phi, \bar{\Phi}] = \int d^4x d^4\theta \, E^{-1} \Omega(\Phi, \bar{\Phi}) + \left[\int d^4x d^2\theta \, \mathcal{E} P(\Phi) + \text{H.c.} \right] \tag{44}$$

in terms of the non-chiral potential $\Omega(\Phi, \bar{\Phi})$ and the chiral superpotential $P(\Phi)$. In our case (43) we find

$$\Omega(\Phi, \bar{\Phi}) = e^{\Phi+\bar{\Phi}} \left[\mathcal{Z}(\Phi) + \bar{\mathcal{Z}}(\bar{\Phi}) \right] , \quad P(\Phi) = -e^{3\Phi} V(\mathcal{Z}(\Phi)) \tag{45}$$

The *Kähler* potential $K(\Phi, \bar{\Phi})$ is given by

$$K = -3\ln(-\frac{\Omega}{3}) \quad \text{or} \quad \Omega = -3e^{-K/3} \tag{46}$$

so that the action (44) is invariant under the supersymmetric (local) Kähler-Weyl transformations

$$K(\Phi, \bar{\Phi}) \to K(\Phi, \bar{\Phi}) + \Lambda(\Phi) + \bar{\Lambda}(\bar{\Phi}) , \qquad P(\Phi) \to -e^{-\Lambda(\Phi)} P(\Phi) \tag{47}$$

with the chiral superfield parameter $\Lambda(\Phi)$. It follows that

$$\mathcal{E} \to e^{\Lambda(\Phi)} \mathcal{E} \tag{48}$$

The scalar potential in terms of the usual fields is given by the standard formula (45)

$$\mathcal{V}(\phi, \bar{\phi}) = e^K \left\{ \left| \frac{\partial P}{\partial \Phi} + \frac{\partial K}{\partial \Phi} P \right|^2 - 3 \left| P \right|^2 \right\} \Bigg| \tag{49}$$

where all the superfields are restricted to their leading field components, $\Phi| = \phi(x)$, and we have introduced the notation

$$\left| \frac{\partial P}{\partial \Phi} + \frac{\partial K}{\partial \Phi} P \right|^2 \equiv |D_\Phi P|^2 = D_\Phi P (K_{\Phi\bar{\Phi}}^{-1}) \bar{D}_{\bar{\Phi}} \bar{P} \tag{50}$$

with $K_{\Phi\bar{\Phi}} = \partial^2 K / \partial\Phi \partial\bar{\Phi}$. Equation (49) can be simplified by making use of the Kähler-Weyl invariance (47) that allows one to choose a gauge

$$P = 1 \tag{51}$$

It is equivalent to the well known fact that the scalar potential (49) is actually governed by the *single* (Kähler-Weyl-invariant) potential

$$G(\Phi, \bar{\Phi}) = \Omega + \ln |P|^2 \tag{52}$$

In our case (45) we find

$$G = e^{\Phi + \bar{\Phi}} \left[\mathcal{Z}(\Phi) + \bar{\mathcal{Z}}(\bar{\Phi}) \right] + 3(\Phi + \bar{\Phi}) + \ln(V(\mathcal{Z}(\Phi)) + \ln(\bar{V}(\bar{\mathcal{Z}}(\bar{\Phi})) \tag{53}$$

So let us choose a gauge by the condition

$$3\Phi + \ln(V(\mathcal{Z}(\Phi)) = 0 \quad \text{or} \quad V(\mathcal{Z}(\Phi)) = e^{-3\Phi} \tag{54}$$

that is equivalent to eq. (51). Then the G-potential (53) gets simplified to

$$G = e^{\Phi + \bar{\Phi}} \left[\mathcal{Z}(\Phi) + \bar{\mathcal{Z}}(\bar{\Phi}) \right] \tag{55}$$

There is the correspondence between a holomorphic function $F(\mathcal{R})$ in the supergravity action (34) and a holomorphic function $\mathcal{Z}(\Phi)$ defining the scalar potential (49),

$$\mathcal{V} = e^G \left[\left(\frac{\partial^2 G}{\partial\Phi \partial\bar{\Phi}} \right)^{-1} \frac{\partial G}{\partial\Phi} \frac{\partial G}{\partial\bar{\Phi}} - 3 \right] \Big| \tag{56}$$

in the classically equivalent scalar-tensor supergravity.

To the end of this section, I would like to comment on the standard way of the inflationary model building by a choice of $K(\Phi, \bar{\Phi})$ and $P(\Phi)$ — see eg., ref. (46) for a recent review.

The factor $\exp(K/M_{\text{Pl}}^2)$ in the F-type scalar potential (49) of the chiral matter-coupled supergravity, in the case of the *canonical* Kähler potential, $K \propto \overline{\Phi}\Phi$, results in the scalar potential $V \propto \exp(|\Phi|^2 / M_{\text{Pl}}^2)$ that is too steep to support chaotic inflation. Actually, it also implies $\eta \approx 1$ or, equivalently, $M_{\text{inflaton}}^2 \approx V_0 / M_{\text{Pl}}^2 \approx H^2$. It is known as the *η-problem* in supergravity (47).

As is clear from our discussion above, the η-problem is not really a supergravity problem, but it is the problem associated with the choice of the canonical Kähler potential for an inflaton superfield. The Kähler potential in supergravity is a (Kähler) gauge-dependent quantity, and its quantum renormalization is not under control. Unlike the one-field inflationary models, a generic Kähler potential is a function of at least two fields, so it implies a nonvanishing curvature in the target space of the *non-linear sigma-model* associated with the Kähler kinetic

term. [3] Hence, a generic Kähler potential cannot be brought to the canonical form by a field redefinion.

To solve the η-problem associated with the simplest (naive) choice of the Kähler potential, on may assume that the Kähler potential K possesses some shift symmetries (leading to its *flat* directions), and then choose inflaton in one such flat direction (49). However, in order to get inflation that way, one also has to add ("by hand") the proper inflaton superpotential breaking the initially introduced shift symmetry, and then stabilize the inflationary trajectory with the help of yet another matter superfield.

The possible alternative is the *D*-term mechanism (50), where inflation is generated in the matter *gauge* sector and, as a result, is highly sensitive to the gauge charges.

It is worth mentioning that in the (perturbative) superstring cosmology one gets the Kähler potential (see e.g., refs. (51; 52))

$$K \propto \log(\text{moduli polynomial})_{\text{CY}} \tag{57}$$

over a Calabi-Yau (CY) space in the type-IIB superstring compactification, thus avoiding the η-problem but leading to a plenty of choices (embarrassment of riches!) in the String Landscape.

Finally, one still has to accomplish stability of a given inflationary model in supergravity against quantum corrections. Such corrections can easily spoil the flatness of the inflaton potential. The Kähler kinetic term is not protected against quantum corrections, because it is given by a full superspace integral (unlike the chiral superpotential term). The $F(\mathcal{R})$ supergravity action (34) is given by a *chiral* superspace integral, so that it is protected against the quantum corrections given by full superspace integrals.

To conclude this section, we claim that an $N = 1$ locally supersymmetric extension of $f(R)$ gravity is possible. It is non-trivial because the auxiliary freedom has to be preserved. The new supergravity action (34) is classically equivalent to the standard $N = 1$ Poincaré supergravity coupled to a *dynamical* chiral matter superfield, whose Kähler potential and the superpotential are dictated by a single *holomorphic* function. Inflaton can be identified with the real scalar field component of that chiral matter superfield originating from the supervielbein.

It is worth noticing that the action (34) allows a natural extension in chiral curved superspace, due to the last equation (31), namely,

$$S_{\text{ext}} = \int d^4x d^2\theta \, \mathcal{E} F(\mathcal{R}, \mathcal{W}^2) + \text{H.c.} \tag{58}$$

where $\mathcal{W}_{\alpha\beta\gamma}$ is the $N = 1$ covariantly-chiral Weyl superfield of the $N = 1$ superspace supergravity, and $\mathcal{W}^2 = \mathcal{W}_{\alpha\beta\gamma}\mathcal{W}^{\alpha\beta\gamma}$. The action (58) also has the auxiliary freedom. In Supersring Theory, the Weyl-tensor-dependence of the gravitational effective action is unambigously determined by the superstring scattering amplitudes or by the super-Weyl invariance of the corresponding non-linear sigma-model (see eg., ref. (48)).

A possible connection of $F(\mathcal{R})$ supergravity to the *Loop Quantum Gravity* was investigated in ref. (3).

---

[3] See eg., ref. (48) for more about the non-linear sigma-models.

7. No-scale $F(\mathcal{R})$ supergravity

In this section we would like to investigate a possibility of spontaneous supersymmetry breaking, without fine tuning, by imposing the condition of the vanishing scalar potential. Those no-scale supergravities are the starting point of many phenomenological applications of supergravity in HEP and inflationary theory, including string theory applications — see eg., refs. (53; 54) and references therein.

The no-scale supergravity arises by demanding the scalar potential (49) to vanish. It results in the vanishing cosmological constant without fine-tuning (55). The no-scale supergravity potential G has to obey the non-linear 2nd-order partial differential equation, which follows from eq. (56),

$$3\frac{\partial^2 G}{\partial\Phi\partial\bar\Phi} = \frac{\partial G}{\partial\Phi}\frac{\partial G}{\partial\bar\Phi} \tag{59}$$

A gravitino mass $m_{3/2}$ is given by the vacuum expectation value (39)

$$m_{3/2} = \left\langle e^{G/2} \right\rangle \tag{60}$$

so that the spontaneous supersymmetry breaking scale can be chosen at will.

The well known exact solution to eq. (59) is given by

$$G = -3\log(\Phi + \bar\Phi) \tag{61}$$

In the recent literature, the no-scale solution (61) is usually modified by other terms, in order to achieve the universe with a positive cosmological constant — see e.g., the KKLT mechanism (56).

To appreciate the difference between the standard no-scale supergravity solution and our 'modified' supergravity, it is worth noticing that the Ansatz (61) is not favoured by our potential (55). In our case, demanding eq. (59) gives rise to the 1st-order non-linear partial differential equation

$$3\left(e^{\bar\Phi}X' + e^{\Phi}\bar X'\right) = \left|e^{\bar\Phi}X' + e^{\Phi}\bar X\right|^2 \tag{62}$$

where we have introduced the notation

$$\mathcal{Z}(\Phi) = e^{-\Phi}X(\Phi), \qquad X' = \frac{dX}{d\Phi} \tag{63}$$

in order to get the differential equation in its most symmetric and concise form.

Accordingly, the gravitino mass (60) is given by

$$m_{3/2} = \left\langle \exp\frac{1}{2}\left(e^{\bar\Phi}X + e^{\Phi}\bar X\right)\right\rangle \tag{64}$$

I am not aware of any non-trivial holomorphic exact solution to eq. (62). However, should it obey a holomorphic differential equation of the form

$$X' = e^{\Phi}g(X,\Phi) \tag{65}$$

with a holomorphic function $g(X,\Phi)$, eq. (62) gives rise to the functional equation

$$3\left(g + \bar g\right) = \left|e^{\bar\Phi}g + \bar X\right|^2 \tag{66}$$

Being restricted to the real variables $\Phi = \bar{\Phi} \equiv y$ and $X = \bar{X} \equiv x$, eq. (62) reads

$$6x' = e^y (x' + x)^2 \,, \quad \text{where} \quad x' = \frac{dx}{dy} \tag{67}$$

This equation can be integrated after a change of variables,

$$x = e^{-y} u \,, \tag{68}$$

and it leads to a quadratic equation with respect to $u' = du/dy$,

$$(u')^2 - 6u' + 6u = 0 \tag{69}$$

It follows

$$y = \int^u \frac{d\xi}{3 \pm \sqrt{3(3 - 2\xi)}} = \mp \sqrt{1 - \tfrac{2}{3}u} + \ln\left(\sqrt{3(3 - 2u)} \pm 3 \right) + C \,. \tag{70}$$

8. Fields from superfields in $F(\mathcal{R})$ supergravity

For simplicity, we set all fermionic fields to zero, when passing to the field components. It greatly simplies most of the field equations, but makes supersymmetry to be manifestly broken (however, SUSY is restored after adding all those fermionic terms back to the action). Applying the standard superspace chiral density formula (38–40)

$$\int d^4 x d^2 \theta \, \mathcal{E} \mathcal{L} = \int d^4 x \, e \left\{ \mathcal{L}_{\text{last}} + B\mathcal{L}_{\text{first}} \right\} \tag{71}$$

to the action (34) yields its bosonic part in the form

$$(-g)^{-1/2} L_{\text{bos}} \equiv f(R, \tilde{R}; X, \bar{X}) = F'(\bar{X}) \left[\tfrac{1}{3} R_* + 4\bar{X}X \right] + 3XF(\bar{X}) + \text{H.c.} \tag{72}$$

where the primes denote differentiation with respect to a given argument. We have used the notation

$$X = \tfrac{1}{3} B \quad \text{and} \quad R_* = R + \frac{i}{2} \varepsilon^{abcd} R_{abcd} \equiv R + i\tilde{R} \tag{73}$$

The $\tilde{R}$ does not vanish in $F(\mathcal{R})$ supergravity, and it represents the *axion* field that is the pseudo-scalar superpartner of real scalaron field in our construction.
Varying eq. (72) with respect to the auxiliary fields X and $\bar{X}$,

$$\frac{\partial L_{\text{bos}}}{\partial X} = \frac{\partial L_{\text{bos}}}{\partial \bar{X}} = 0 \tag{74}$$

gives rise to the algebraic equations on the auxiliary fields,

$$3\bar{F} + X(4\bar{F}' + 7F') + 4\bar{X}XF'' + \tfrac{1}{3}F'' R_* = 0 \tag{75}$$

and its conjugate

$$3F + \bar{X}(4F' + 7\bar{F}') + 4\bar{X}X\bar{F}'' + \tfrac{1}{3}\bar{F}'' \bar{R}_* = 0 \tag{76}$$

where $F = F(X)$ and $\bar{F} = \bar{F}(\bar{X})$. The algebraic equations (75) and (76) cannot be explicitly solved for X in a generic $F(\mathcal{R})$ supergravity.

To recover the standard (pure) supergravity in our approach, let us consider the simple special case when

$$F'' = 0 \qquad \text{or, equivalently,} \qquad F(\mathcal{R}) = f_0 - \tfrac{1}{2}f_1\mathcal{R} \tag{77}$$

with some complex constants f_0 and f_1, where $\text{Re}\,f_1 > 0$. Then eq. (75) is easily solved as

$$\bar{X} = \frac{3f_0}{5(\text{Re}\,f_1)} \tag{78}$$

Substituting this solution back into the Lagrangian (72) yields

$$L = -\tfrac{1}{3}(\text{Re}\,f_1)R + \frac{9\,|f_0|^2}{5(\text{Re}\,f_1)} \equiv -\tfrac{1}{2}M_{\text{Pl}}^2 R - \Lambda \tag{79}$$

where we have introduced the reduced Planck mass M_{Pl}, and the cosmological constant Λ as

$$\text{Re}\,f_1 = \tfrac{3}{2}M_{\text{Pl}}^2 \qquad \text{and} \qquad \Lambda = \frac{-6\,|f_0|^2}{5M_{\text{Pl}}^2} \tag{80}$$

It is the standard pure supergravity with a *negative* cosmological constant (38–40).

9. Generic $\mathcal{R}^2$ supergravity, and AdS bound

The simplest non-trivial $F(\mathcal{R})$ supergravity is obtained by choosing $F'' = const. \neq 0$ that leads to the $\mathcal{R}^2$-supergravity defined by a generic *quadratic* polynomial in terms of the scalar supercurvature (8).

Let us recall that the stability conditions in $f(R)$-gravity are given by eqs. (12) in the notation (9). In the notation (72) used here, ie. when $f(R) = -\tfrac{1}{2}M_{\text{Pl}}^2\tilde{f}(R)$, one gets the opposite signs,

$$f'(R) < 0 \tag{81}$$

and

$$f''(R) > 0 \tag{82}$$

The first (classical stability) condition (81) is related to the sign factor in front of the Einstein-Hilbert term (linear in R) in the $f(R)$-gravity action, and it ensures that graviton is not a ghost. The second (quantum stability) condition (82) guarantees that scalaron is not a tachyon.

Being mainly interested in the inflaton part of the bosonic $f(R)$-gravity action that follows from eq. (72), we set both gravitino and axion to zero, which also implies $R_* = R$ and a *real X*. In $F(R)$ supergravity the stability condition (81) is to be replaced by a stronger condition,

$$F'(X) < 0 \tag{83}$$

It is easy to verify that eq. (81) follows from eq. (83) because of eq. (74). Equation (83) also ensures the classical stability of the bosonic $f(R)$ gravity embedding into the full $F(\mathcal{R})$ supergravity against small fluctuations of the axion field.

Let's now investigate the most general non-trivial Ansatz (with $F'' = const. \neq 0$) for the $F(R)$ supergravity function in the form

$$F(\mathcal{R}) = f_0 - \frac{1}{2}f_1\mathcal{R} + \frac{1}{2}f_2\mathcal{R}^2 \tag{84}$$

with three coupling constants f_0, f_1 and f_2. We will take all of them to be real, since we will ignore this potential source of *CP*-violation here (see, however, the Outlook). As regards the mass dimensions of the quantities introduced, we have

$$[F] = [f_0] = 3, \quad [R] = [f_1] = 2, \quad \text{and} \quad [\mathcal{R}] = [f_2] = 1 \tag{85}$$

The bosonic Lagrangian (72) with the function (84) reads

$$(-g)^{-1/2} L_{\text{bos}} = 11 f_2 X^3 - 7 f_1 X^2 + \left(\tfrac{2}{3} f_2 R + 6 f_0 \right) X - \tfrac{1}{3} f_1 R \tag{86}$$

Hence, the auxiliary field equation (74) takes the form of a *quadratic* equation,

$$\tfrac{33}{2} f_2 X^2 - 7 f_1 X + \tfrac{1}{3} R f_2 + 3 f_0 = 0 \tag{87}$$

whose solution is given by

$$X_\pm = \frac{7}{3 \cdot 11} \left[\frac{f_1}{f_2} \pm \sqrt{ \frac{2 \cdot 11}{7^2} (R_{\text{max}} - R) } \right] \tag{88}$$

where we have introduced the *maximal* scalar curvature

$$R_{\text{max}} = \frac{7^2}{2 \cdot 11} \frac{f_1^2}{f_2^2} - 3^2 \frac{f_0}{f_2} \tag{89}$$

Equation (88) obviously implies the automatic bound on the scalar curvature (from one side only). In our notation, it corresponds to the (AdS) bound on the scalar curvature from above,

$$R < R_{\text{max}} \tag{90}$$

The existence of the built-in maximal (upper) scalar curvature (or the AdS bound) is a nice bonus of our construction. It is similar to the factor $\sqrt{1 - v^2/c^2}$ in Special Relativity. Yet another close analogy comes from the *Born-Infeld* non-linear extension of Maxwell electrodynamics, whose (dual) Hamiltonian is proportional to (48)

$$\left(1 - \sqrt{ 1 - \vec{E}^2 / E_{\text{max}}^2 - \vec{H}^2 / H_{\text{max}}^2 + (\vec{E} \times \vec{H})^2 / E_{\text{max}}^2 H_{\text{max}}^2 } \right) \tag{91}$$

in terms of the electric and magnetic fields $\vec{E}$ and $\vec{H}$, respectively, with their maximal values. For instance, in String Theory one has $E_{\text{max}} = H_{\text{max}} = (2\pi\alpha')^{-1}$ (48).

Substituting the solution (88) back into eq. (86) yields the corresponding $f(R)$-gravity Lagrangian

$$f_\pm(R) = \frac{2 \cdot 7}{11} \frac{f_0 f_1}{f_2} - \frac{2 \cdot 7^3}{3^3 \cdot 11^2} \frac{f_1^3}{f_2^2}$$
$$- \frac{19}{3^2 \cdot 11} f_1 R \mp \sqrt{\frac{2}{11}} \left(\frac{2^2}{3^3} f_2 \right) (R_{\text{max}} - R)^{3/2} \tag{92}$$

Expanding eq. (92) into power series of R yields

$$f_\pm(R) = -\Lambda_\pm - a_\pm R + b_\pm R^2 + \mathcal{O}(R^3) \tag{93}$$

whose coefficients are given by

$$\Lambda_\pm = \frac{2 \cdot 7}{3^2 \cdot 11} f_1 \left(R_{max} - \frac{7^2}{2 \cdot 3 \cdot 11} \frac{f_1^2}{f_2^2} \right) \pm \sqrt{\frac{2}{11}} \left(\frac{2^2}{3^3} f_2 \right) R_{max}^{3/2} \tag{94}$$

$$a_\pm = \frac{19}{3^2 \cdot 11} f_1 \mp \sqrt{\frac{2}{11} R_{max}} \left(\frac{2}{3^2} f_2 \right) \tag{95}$$

and

$$b_\pm = \mp \sqrt{\frac{2}{11 R_{max}}} \left(\frac{f_2}{2 \cdot 3^2} \right) \tag{96}$$

Those equations greatly simplify when $f_0 = 0$. One finds (5; 8)

$$f_\pm^{(0)}(R) = \frac{-5 \cdot 17 M_{Pl}^2}{2 \cdot 3^2 \cdot 11} R + \frac{2 \cdot 7}{3^2 \cdot 11} M_{Pl}^2 (R - R_{max}) \left[1 \pm \sqrt{1 - R/R_{max}} \right] \tag{97}$$

where we have chosen

$$f_1 = \frac{3}{2} M_{Pl}^2 \tag{98}$$

in order to get the standard normalization of the Einstein-Hilbert term that is linear in R. Then, in the limit $R_{max} \to +\infty$, both functions $f_\pm^{(0)}(R)$ reproduce General Relativity. In another limit $R \to 0$, one finds a *vanishing* or *positive* cosmological constant,

$$\Lambda_-^{(0)} = 0 \qquad \text{and} \qquad \Lambda_+^{(0)} = \frac{2^2 \cdot 7}{3^2 \cdot 11} M_{Pl}^2 R_{max} \tag{99}$$

The stability conditions are given by eqs. (81), (82) and (83), while the 3rd condition implies the 2nd one. In our case (92) we have

$$f_\pm'(R) = -\frac{19}{3^2 \cdot 11} f_1 \pm \sqrt{\frac{2}{11}} \left(\frac{2}{3^2} f_2 \right) \sqrt{R_{max} - R} \ < 0 \tag{100}$$

and

$$f_\pm''(R) = \mp \left(\frac{f_2}{3^2} \right) \sqrt{\frac{2}{11(R_{max} - R)}} \ > 0 \tag{101}$$

while eqs. (83), (84) and (88) yield

$$\pm \sqrt{\frac{2 \cdot 11}{7^2} (R_{max} - R)} \ < \frac{19}{2 \cdot 7} \frac{f_1}{f_2} \tag{102}$$

It follows from eq. (101) that

$$f_2^{(+)} < 0 \qquad \text{and} \qquad f_2^{(-)} > 0 \tag{103}$$

Then the stability condition (82) is obeyed for any value of R.
As regards the $(-)$-case, there are *two* possibilities depending upon the sign of f_1. Should f_1 be *positive*, all the remaining stability conditions are automatically satisfied, ie. in the case of both $f_2^{(-)} > 0$ and $f_1^{(-)} > 0$.

Should f_1 be *negative*, $f_1^{(-)} < 0$, we find that the remaining stability conditions (100) and (102) are *the same*, as they should, while they are both given by

$$R < R_{\max} - \frac{19^2}{2^3 \cdot 11} \frac{f_1^2}{f_2^2} = -\frac{3 \cdot 5}{2^3 \cdot 11} \frac{f_1^2}{f_2^2} - 3^2 \frac{f_0}{f_2} \equiv R_{\max}^{\text{ins}} \tag{104}$$

As regards the $(+)$-case, eq. (102) implies that f_1 should be *negative*, $f_1 < 0$, whereas then eqs. (100) and (102) result in *the same* condition (104) again.

Since $R_{\max}^{\text{ins}} < R_{\max}$, our results imply that the instability happens *before* R reaches $R_{\max}$ in all cases with negative f_1.

As regards the particularly simple case (97), the stability conditions allow us to choose the lower sign only.

A different example arises with a negative f_1. When choosing the lower sign (ie. a positive f_2) for definiteness, we find

$$\begin{aligned}
f_-(R) = &-\frac{2 \cdot 7}{11} f_0 \left| \frac{f_1}{f_2} \right| + \frac{2 \cdot 7^3}{3^3 \cdot 11^2} \left| \frac{f_1^3}{f_2^2} \right| \\
&+ \frac{19}{3^2 \cdot 11} |f_1| R + \sqrt{\frac{2}{11}} \left(\frac{2^2}{3^3} f_2 \right) (R_{\max} - R)^{3/2}
\end{aligned} \tag{105}$$

Demanding the standard normalization of the Einstein-Hilbert term in this case implies

$$R_{\max} = \frac{3^4 \cdot 11}{2^3 f_2^2} \left(\frac{M_{\text{Pl}}^2}{2} + \frac{19}{3^2 \cdot 11} |f_1| \right)^2 \tag{106}$$

where we have used eq. (95). It is easy to verify by using eq. (94) that the cosmological constant is always *negative* in this case, and the instability bound (104) is given by

$$R_{\max}^{\text{ins}} = \frac{3^4 \cdot 11 M_{\text{Pl}}^2}{2^3 f_2^2} \left(\frac{M_{\text{Pl}}^2}{2^2} + \frac{19 |f_1|}{3^2 \cdot 11} \right) < R_{\max} \tag{107}$$

The $f_-(R)$ function of eq. (92) can be rewritten to the form

$$f(R) = \frac{7^3}{3^3 \cdot 11^2} \frac{f_1^3}{f_2^2} - \frac{2 \cdot 7}{3^2 \cdot 11} f_1 R_{\max} - \frac{19}{3^2 \cdot 11} f_1 R + f_2 \sqrt{\frac{2^5}{3^6 \cdot 11}} (R_{\max} - R)^{3/2} \tag{108}$$

where we have used eq. (89). There are *three* physically different regimes:

(i) the *high-curvature regime*, $R < 0$ and $|R| \gg R_{\max}$. Then eq. (108) implies

$$f(R) \approx -\Lambda_h - a_h R + c_h |R|^{3/2} \tag{109}$$

whose coefficients are given by

$$\begin{aligned}
\Lambda_h &= \frac{2 \cdot 7}{3^2 \cdot 11} f_1 R_{\max} - \frac{7^3}{3^3 \cdot 11^2} \frac{f_1^3}{f_2^2} \quad, \\
a_h &= \frac{19}{3^2 \cdot 11} f_1 \quad, \\
c_h &= \sqrt{\frac{2}{11}} \left(\frac{2^2}{3^3} f_2 \right)
\end{aligned} \tag{110}$$

(ii) the *low-curvature regime*, $|R/R_{\max}| \ll 1$. Then eq. (108) implies

$$f(R) \approx -\Lambda_l - a_l R \ , \tag{111}$$

whose coefficients are given by

$$
\begin{aligned}
\Lambda_l &= \Lambda_h - \sqrt{\frac{2R_{\max}^3}{11}} \left(\frac{2^2}{3^3} f_2 \right) \ , \\[2mm]
a_l &= a_h + \sqrt{\frac{2R_{\max}}{11}} \left(\frac{2}{3^2} f_2 \right) = a_- = \frac{M_{\mathrm{Pl}}^2}{2} \ ,
\end{aligned}
\tag{112}
$$

where we have used eq. (95).

(iii) the *near-the-bound regime* (assuming that no instability happens before it), $R = R_{\max} + \delta R$, $\delta R < 0$, and $|\delta R / R_{\max}| \ll 1$. Then eq. (108) implies

$$f(R) \approx -\Lambda_b + a_b \, |\delta R| + c_b \, |\delta R|^{3/2} \tag{113}$$

whose coefficients are

$$
\begin{aligned}
\Lambda_b &= \frac{1}{3} f_1 R_{\max} - \frac{7^3}{3^3 \cdot 11^2} \frac{f_1^3}{f_2^2} \ , \\[2mm]
a_b &= a_h \ , \\[2mm]
c_b &= \sqrt{\frac{2}{11}} \left(\frac{2^2}{3^3} f_2 \right)
\end{aligned}
\tag{114}
$$

The cosmological dynamics may be either directly derived from the gravitational equations of motion in the $f(R)$-gravity with a given function $f(R)$, or just read off from the form of the corresponding scalar potential of a scalaron (see below). For instance, as was demonstrated in ref. (5) for the special case $f_0 = 0$, a cosmological expansion is possible in the regime (i) towards the regime (ii), and then, perhaps, to the regime (iii) unless an instability occurs.

However, one should be careful since our toy-model (84) does not pretend to be viable in the low-curvature regime, eg., for the present Universe. Nevertheless, if one wants to give it some physical meaning there, by identifying it with General Relativity, then one should also fine-tune the cosmological constant Λ_l in eq. (112) to be "small" and positive. We find that it amounts to

$$R_{\max} \approx \frac{3^4 \cdot 7^2 \cdot 11}{2^5 \cdot 19^2} \frac{M_{\mathrm{Pl}}^4}{f_2^2} \equiv R_{\Lambda=0} \tag{115}$$

with the actual value of $R_{\max}$ to be "slightly" above of that bound, $R_{\max} > R_{\Lambda=0}$. It is also posssible to have the vanishing cosmological constant, $\Lambda_l = 0$, when choosing $R_{\max} = R_{\Lambda=0}$. It is worth mentioning that it relates the values of $R_{\max}$ and f_2.

The particular $\mathcal{R}^2$-supergravity model (with $f_0 = 0$) was introduced in ref. (5) in an attempt to get viable embedding of the Starobinsky model into $F(\mathcal{R})$-supergravity. However, it failed because, as was found in ref. (5), the higher-order curvature terms cannot be ignored in eq. (97), ie. the R^n-terms with $n \geq 3$ are not small enough against the R^2-term. In fact, the possibility of destabilizing the Starobinsky inflationary scenario by the terms with higher powers of the scalar curvature, in the context of $f(R)$ gravity, was noticed earlier in refs. (57; 58). The most general Ansatz (84), which is merely *quadratic* in the supercurvature, does not help for that purpose either.

For example, the full $f(R)$-gravity function $f_-(R)$ in eq. (97), which we derived from our $\mathcal{R}^2$-supergravity, gives rise to the inflaton scalar potential

$$V(y) = V_0 \left(11e^y + 3\right) \left(e^{-y} - 1\right)^2 \tag{116}$$

where $V_0 = (3^3/2^6) M_{\text{Pl}}^4 / f_2^2$. The corresponding inflationary parameters

$$\varepsilon(y) = \frac{1}{3} \left[\frac{e^y \left(11 + 11e^{-y} + 6e^{-2y}\right)}{(11e^y + 3)(e^{-y} - 1)} \right]^2 \geq \frac{1}{3} \tag{117}$$

and

$$\eta(y) = \frac{2}{3} \frac{\left(11e^y + 5e^{-y} + 12e^{-2y}\right)}{(11e^y + 3)(e^{-y} - 1)^2} \geq \frac{2}{3} \tag{118}$$

are not small enough for matching the WMAP observational data. A solution to this problem is given in the next Sec. 10.

10. Chaotic inflation in $F(\mathcal{R})$ supergravity

Let us take now one more step further and consider a new Ansatz for $F(\mathcal{R})$ function in the *cubic* form

$$F(\mathcal{R}) = -\tfrac{1}{2} f_1 \mathcal{R} + \tfrac{1}{2} f_2 \mathcal{R}^2 - \tfrac{1}{6} f_3 \mathcal{R}^3 \tag{119}$$

whose real (positive) coupling constants $f_{1,2,3}$ are of (mass) dimension 2, 1 and 0, respectively. Our conditions on the coefficients are

$$f_3 \gg 1 \,, \qquad f_2^2 \gg f_1 \tag{120}$$

The first condition is needed to have inflation at the curvatures much less than M_{Pl}^2 (and to meet observations), while the second condition is needed to have the scalaron (inflaton) mass be much less than M_{Pl}, in order to avoid large (gravitational) quantum loop corrections after the end of inflation up to the present time.

The bosonic action is given by eq. (72). In the case of a real scalaron it reduces to

$$L/\sqrt{-g} = 2F' \left[\frac{1}{3} R + 4X^2 \right] + 6XF \tag{121}$$

so that the real auxiliary field is a solution to the algebraic equation

$$3F + 11F'X + F'' \left[\frac{1}{3} R + 4X^2 \right] = 0 \tag{122}$$

Stability of the bosonic embedding in supergravity requires $F'(X) < 0$ (Sec. 9). In the case (119) it gives rise to the condition $f_2^2 < f_1 f_3$. For simplicity here, we will assume a stronger condition,

$$f_2^2 \ll f_1 f_3 \tag{123}$$

Then the second term on the right-hand-side of eq. (119) will not affect inflation, as is shown below.

Equation (121) with the Ansatz (119) reads

$$L = -5f_3 X^4 + 11f_2 X^3 - (7f_1 + \tfrac{1}{3} f_3 R)X^2 + \tfrac{2}{3} f_2 RX - \tfrac{1}{3} f_1 R \tag{124}$$

and gives rise to a cubic equation on X,

$$X^3 - \left(\frac{33f_2}{20f_3}\right) X^2 + \left(\frac{7f_1}{10f_3} + \frac{1}{30}R\right) X - \frac{f_2}{30f_3}R = 0 \tag{125}$$

We find three consecutive (overlapping) regimes.

- The high curvature regime including inflation is given by

$$\delta R < 0 \quad \text{and} \quad \frac{|\delta R|}{R_0} \gg \left(\frac{f_2^2}{f_1 f_3}\right)^{1/3} \tag{126}$$

where we have introduced the notation $R_0 = 21f_1/f_3 > 0$ and $\delta R = R + R_0$. With our sign conventions we have $R < 0$ during the de Sitter and matter dominated stages. In the regime (126) the f_2-dependent terms in eqs. (124) and (125) can be neglected, and we get

$$X^2 = -\frac{1}{30}\delta R \tag{127}$$

and

$$L = -\frac{f_1}{3}R + \frac{f_3}{180}(R + R_0)^2 \tag{128}$$

It closely reproduces the Starobinsly inflationary model (Sec. 2) since inflation occurs at $|R| \gg R_0$. In particular, we can identify

$$f_3 = \frac{15M_{\text{Pl}}^2}{M_{\text{inf}}^2} \tag{129}$$

It is worth mentioning that we cannot simply set $f_2 = 0$ in eq. (119) because it would imply $X = 0$ and $L = -\frac{f_1}{3}R$ for $\delta R > 0$. As a result of that the scalar degree of freedom would disappear that would lead to the breaking of a regular Cauchy evolution. Therefore, the second term in eq. (119) is needed to remove that degeneracy.

- The intermediate (post-inflationary) regime is given by

$$\frac{|\delta R|}{R_0} \ll 1 \tag{130}$$

In this case X is given by a root of the cubic equation

$$30X^3 + (\delta R)X + \frac{f_2 R_0}{f_3} = 0 \tag{131}$$

It also implies that the 2nd term in eq. (125) is always small. Equation (131) reduces to eq. (127) under the conditions (126).

- The low-curvature regime (up to $R = 0$) is given by

$$\delta R > 0 \quad \text{and} \quad \frac{\delta R}{R_0} \gg \left(\frac{f_2^2}{f_1 f_3}\right)^{1/3} \tag{132}$$

It yields

$$X = \frac{f_2 R}{f_3(R + R_0)} \tag{133}$$

and

$$L = -\frac{f_1}{3}R + \frac{f_2^2 R^2}{3f_3(R + R_0)} \tag{134}$$

It is now clear that f_1 should be equal to $3M_{Pl}^2/2$ in order to obtain the correctly normalized Einstein gravity at $|R| \ll R_0$. In this regime the scalaron mass squared is given by

$$\frac{1}{3|f''(R)|} = \frac{f_3 R_0 M_{Pl}^2}{4f_2^2} = \frac{21f_1}{4f_2^2}M_{Pl}^2 = \frac{63M_{Pl}^4}{8f_2^2} \tag{135}$$

in agreement with the case of the absence of the $\mathcal{R}^3$ term, studied in the previous section. The scalaron mass squared (135) is much less than M_{Pl}^2 indeed, due to the second inequality in eq. (120), but it is much more than one at the end of inflation ($\sim M^2$).

It is worth noticing that the corrections to the Einstein action in eqs. (128) and (134) are of ther same order (and small) at the borders of the intermediate region (130).

The roots of the cubic equation (125) are given by the textbook (Cardano) formula (59), though that formula is not very illuminating in a generic case. The Cardano formula greatly simplifies in the most interesting (high curvature) regime where inflation takes place, and the Cardano discriminant is

$$D \approx \left(\frac{R}{90}\right)^3 < 0 \tag{136}$$

It implies that all three roots are real and unequal. The Cardano formula yields the roots

$$X_{1,2,3} \approx \frac{2}{3}\sqrt{\frac{-R}{10}}\cos\left(\frac{27}{4f_3\sqrt{-10R/f_2^2}} + C_{1,2,3}\right) + \frac{11f_2}{20f_3} \tag{137}$$

where the constant $C_{1,2,3}$ takes the values $(\pi/6, 5\pi/6, 3\pi/2)$.

As regards the leading terms, eqs. (124) and (137) result in the $(-R)^{3/2}$ correction to the $(R + R^2)$-terms in the effective Lagrangian in the high-curvature regime $|R| \gg f_2^2/f_3^2$. In order to verify that this correction does not change our results under the conditions (126), let us consider the $f(R)$-gravity model with

$$\tilde{f}(R) = R - b(-R)^{3/2} - aR^2 \tag{138}$$

whose parameters $a > 0$ and $b > 0$ are subject to the conditions $a \gg 1$ and $b/a^2 \ll 1$. It is easy to check that $\tilde{f}'(R) > 0$ for $R \in (-\infty, 0]$, as is needed for (classical) stability.

Any $f(R)$ gravity model is classically equivalent to the scalar-tensor gravity with certain scalar potential (Sec. 3). The scalar potential can be calculated from a given function $f(R)$ along the standard lines (Sec. 3). We find (in the high curvature regime)

$$V(y) = \frac{1}{8a}\left(1 - e^{-y}\right)^2 + \frac{b}{8\sqrt{2a}}e^{-2y}\left(e^y - 1\right)^{3/2} \tag{139}$$

in terms of the inflaton field y. The first term of this equation is the scalar potential associated with the pure $(R + R^2)$ model, and the 2nd term is the correction due to the $R^{3/2}$-term in eq. (138). It is now clear that for large positive y the vacuum energy in the first term dominates and drives inflation until the vacuum energy is compensated by the y-dependent terms near $e^y = 1$.

It can be verified along the lines of ref. (33) that the formula for scalar perturbations remains the same as that for the model (7), ie. $\Delta_{\mathcal{R}}^2 \approx N^2 M_{\text{inf}}^2/(24\pi^2 M_{\text{Pl}}^2)$, where N is the number of e-folds from the end of inflation. So, to fit the observational data, one has to choose

$$f_3 \approx 5N_e^2/(8\pi^2\Delta_{\mathcal{R}}^2) \approx 6.5 \cdot 10^{10}(N_e/50)^2 \tag{140}$$

Here the value of $\Delta_{\mathcal{R}}$ is taken from ref. (14) and the subscript $\mathcal{R}$ has a different meaning from the rest of this paper.

We conclude that the model (119) with a sufficiently small f_2 obeying the conditions (120) and (123) gives a viable realization of the chaotic $(R + R^2)$-type inflation in supergravity. The only significant difference with respect to the original $(R + R^2)$ inflationary model is the scalaron mass that becomes much larger than M in supergravity, soon after the end of inflation when δR becomes positive. However, it only makes the scalaron decay faster and creation of the usual matter (reheating) more effective.

The whole series in powers of $\mathcal{R}$ may also be considered, instead of the limited Ansatz (119). The only necessary condition for embedding inflation is that f_3 should be anomalously large. When the curvature grows, the $\mathcal{R}^3$-term should become important much earlier than the convergence radius of the whole series without that term. Of course, it means that viable inflation may not occur for any function $F(\mathcal{R})$ but only inside a small region of non-zero measure in the space of all those functions. However, the same is true for all known inflationary models, so the very existence of inflation has to be taken from the observational data, not from a pure thought.

The results of this Section can be considered as the viable alternative to the earlier fundamental proposals (49; 50) for realization of chaotic inflation in supergravity. But inflation is not the only target of our construction. As is well known (20; 21; 60), the scalaron decays into pairs of particles and anti-particles of quantum matter fields, while its decay into gravitons is strongly suppressed (61). It thus represents the *universal mechanism of viable reheating* after inflation and provides a transition to the subsequent hot radiation-dominated stage of the Universe evolution. In its turn, it leads to the standard primordial nucleosynthesis (BBN) after. In $F(R)$ supergravity the scalaron has a pseudo-scalar superpartner (axion) that may be the source of a strong CP-violation and then, subsequently, lepto- and baryo-genesis that naturally lead to baryon (matter-antimatter) asymmetry (65; 66) — see Secs. 12 and 14 for more.

11. Nonminimal scalar-curvature coupling in gravity and supergravity

It was recently proposed in refs. (67; 68; 70) to identify Higgs scalar with inflaton, by employing a nonminimal coupling of the Higgs scalar to the scalar curvature of spacetime. Adding such nonminimal coupling to gravity is natural in curved spacetime because it is required by renormalization (71).

Let us compare the inflationary scalar potential, derived by the use of the nonminimal coupling (67; 68; 70), with the scalar potential that follows from the $(R + R^2)$ inflationary model (Sec. 2), and confirm their equivalence. In what follows we will upgrade that equivalence to supergravity. In this section we set $M_{\text{Pl}} = 1$ too.

The original motivation of refs. (67; 68; 70) was based on the assumption that there is no new physics beyond the Standard Model up to the Planck scale. Then it is natural to search for the Higgs mechanism of inflation by identifying inflaton with Higgs. Our motivation here is different: we assume that there is the new physics beyond the Standard Model, and it is given by supersymmetry. Then it is quite natural to search for most economical mechanisms

of inflation in the context of supergravity. Moreover, we do not have to identify inflaton with a Higgs particle of the Minimal Supersymmetric Standard Model.

Let us begin with the 4D Lagrangian

$$\mathcal{L}_J = \sqrt{-g_J} \left\{ -\frac{1}{2}(1 + \xi\phi_J^2)R_J + \frac{1}{2}g_J^{\mu\nu}\partial_\mu\phi_J\partial_\nu\phi_J - V(\phi_J) \right\} \tag{141}$$

where we have introduced the real scalar field $\phi_J(x)$, nonminimally coupled to gravity (with the coupling constant ξ) in Jordan frame, with the Higgs-like scalar potential

$$V(\phi_J) = \frac{\lambda}{4}(\phi_J^2 - v^2)^2 \tag{142}$$

The action (141) can be rewritten to Einstein frame by redefining the metric via a Weyl transformation,

$$g^{\mu\nu} = \frac{g_J^{\mu\nu}}{(1 + \xi\phi_J^2)} \tag{143}$$

It gives rise to the standard Einstein-Hilbert term $(-\frac{1}{2}R)$ for gravity in the Lagrangian. However, it also leads to a nonminimal (or noncanonical) kinetic term of the scalar field ϕ_J. To get the canonical kinetic term, a scalar field redefinition is needed, $\phi_J \to \varphi(\phi_J)$, subject to the condition

$$\frac{d\varphi}{d\phi_J} = \frac{\sqrt{1 + \xi(1 + 6\xi)\phi_J^2}}{1 + \xi\phi_J^2} \tag{144}$$

As a result, the non-minimal theory (141) is classically equivalent to the standard (canonical) theory of the scalar field $\varphi(x)$ minimally coupled to gravity,

$$\mathcal{L}_E = \sqrt{-g} \left\{ -\frac{1}{2}R + \frac{1}{2}g^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi - V(\varphi) \right\} \tag{145}$$

with the scalar potential

$$V(\varphi) = \frac{V(\phi_J(\varphi))}{[1 + \xi\phi_J^2(\varphi)]^2} \tag{146}$$

Given a large positive $\xi \gg 1$, in the small field limit one finds from eq. (144) that $\phi_J \approx \varphi$, whereas in the large φ l imit one gets

$$\varphi \approx \sqrt{\frac{3}{2}}\log(1 + \xi\phi_J^2) \tag{147}$$

Then eq. (146) yields the scalar potential:

(i) in the *very small* field limit, $\varphi < \sqrt{\frac{2}{3}}\xi^{-1}$, as

$$V_{\text{vs}}(\varphi) \approx \frac{\lambda}{4}\varphi^4 \tag{148}$$

(ii) in the *small* field limit, $\sqrt{\frac{2}{3}}\xi^{-1} < \varphi \ll \sqrt{\frac{3}{2}}$, as

$$V_{\mathrm{s}}(\varphi) \approx \frac{\lambda}{6\xi^2}\varphi^2, \tag{149}$$

(iii) and in the *large* field limit, $\varphi \gg \sqrt{\frac{2}{3}}\xi^{-1}$, as

$$V(\varphi) \approx \frac{\lambda}{4\xi^2}\left(1 - \exp\left[-\sqrt{\frac{2}{3}}\varphi\right]\right)^2 \tag{150}$$

We have assumed here that $\xi \gg 1$ and $v\xi \ll 1$.

Identifying inflaton with Higgs particle requires the parameter v to be of the order of weak scale, and the coupling λ to be the Higgs boson selfcoupling at the inflationary scale. The scalar potential (150) is perfectly suitable to support a slow-roll inflation, while its consistency with the COBE normalization condition (Sec. 4) for the observed CMB amplitude of density perturbations (eg., at the e-foldings number $N_e = 50 \div 60$) gives rise to the relation $\xi/\sqrt{\lambda} \approx \mathcal{O}(10^5)$ (67; 68; 70).

The scalar potential (149) corresponds to the post-inflationary matter-dominated epoch described by the oscillating inflaton field φ with the frequency

$$\omega = \sqrt{\frac{\lambda}{3}}\,\xi^{-1} = M_{\mathrm{inf}} \tag{151}$$

When gravity is extended to 4D, $N = 1$ supergravity, any physical real scalar field should be complexified, becoming the leading complex scalar field component of a chiral (scalar) matter supermultiplet. In a curved superspace of $N = 1$ supergravity, the chiral matter supermultiplet is described by a covariantly chiral superfield Φ obeying the constaraint $\overline{\nabla}_{\dot{\alpha}}\Phi = 0$. The standard (generic and minimally coupled) matter-supergravity action is given by in superspace by eqs. (44) and and (46), namely,

$$S_{\mathrm{MSG}} = -3\int d^4x d^4\theta E^{-1}\exp\left[-\frac{1}{3}K(\Phi,\overline{\Phi})\right] + \left\{\int d^4x d^2\theta \mathcal{E}W(\Phi) + \text{H.c.}\right\} \tag{152}$$

in terms of the Kähler potential $K = -3\log(-\frac{1}{3}\Omega)$ and the superpotential W of the chiral supermatter, and the full density E and the chiral density $\mathcal{E}$ of the superspace supergravity. The non-minimal matter-supergravity coupling in superspace reads

$$S_{\mathrm{NM}} = \int d^4x d^2\theta \mathcal{E}X(\Phi)\mathcal{R} + \text{H.c.} \tag{153}$$

in terms of the chiral function $X(\Phi)$ and the N=1 chiral scalar supercurvature superfield $\mathcal{R}$ obeying $\overline{\nabla}_{\dot{\alpha}}\mathcal{R} = 0$. In terms of the field components of the superfields the non-minimal action (153) is given by

$$\int d^4x d^2\theta \mathcal{E}X(\Phi)\mathcal{R} + \text{H.c.} = -\frac{1}{6}\int d^4x\sqrt{-g}X(\phi_c)R + \text{H.c.} + \dots \tag{154}$$

stand for the fermionic terms, and $\phi_c = \Phi| = \phi + i\chi$ is the leading complex scalar field component of the superfield Φ. Given $X(\Phi) = -\xi\Phi^2$ with the real coupling constant ξ, we find the bosonic contribution

$$S_{\text{NM,bos.}} = \frac{1}{6}\xi \int d^4x \sqrt{-g} \left(\phi^2 - \chi^2 \right) R \tag{155}$$

It is worth noticing that the supersymmetrizable (bosonic) non-minimal coupling reads $\left[\phi_c^2 + (\phi_c^\dagger)^2 \right] R$, not $(\phi_c^\dagger \phi_c)R$.

Let us now introduce the manifestly supersymmetric nonminimal action (in Jordan frame) as

$$S = S_{\text{MSG}} + S_{\text{NM}} \tag{156}$$

In curved superspace of $N = 1$ supergravity the (Siegel's) chiral integration rule

$$\int d^4x d^2\theta \mathcal{E} \mathcal{L}_{\text{ch}} = \int d^4x d^4\theta E^{-1} \frac{\mathcal{L}_{\text{ch}}}{\mathcal{R}} \tag{157}$$

applies to any chiral superfield Lagrangian $\mathcal{L}_{\text{ch}}$ with $\overline{\nabla}_{\dot{\alpha}} \mathcal{L}_{\text{ch}} = 0$. It is, therefore, possible to rewrite eq. (153) to the equivalent form

$$S_{\text{NM}} = \int d^4x d^4\theta E^{-1} \left[X(\Phi) + \overline{X}(\overline{\Phi}) \right] \tag{158}$$

We conclude that adding S_{NM} to S_{MSG} is equivalent to the simple change of the Ω-potential as (*cf.* ref. (72))

$$\Omega \rightarrow \Omega_{\text{NM}} = \Omega + X(\Phi) + \bar{X}(\overline{\Phi}) \tag{159}$$

It amounts to the change of the Kähler potential as

$$K_{\text{NM}} = -3\ln \left[e^{-K/3} - \frac{X(\Phi) + \overline{X}(\overline{\Phi})}{3} \right] \tag{160}$$

The scalar potential in the matter-coupled supergravity (152) is given by eq. (56),

$$V(\phi, \bar{\phi}) = e^G \left[\left(\frac{\partial^2 G}{\partial\phi\partial\bar{\phi}} \right)^{-1} \frac{\partial G}{\partial\phi} \frac{\partial G}{\partial\bar{\phi}} - 3 \right] \tag{161}$$

in terms of the Kähler-gauge-invariant function (52), ie.

$$G = K + \ln |W|^2 \tag{162}$$

Hence, in the nonminimal case (156) we have

$$G_{\text{NM}} = K_{\text{NM}} + \ln |W|^2 \tag{163}$$

Contrary to the bosonic case, one gets a nontrivial Kähler potential K_{NM}, ie. a *Non-Linear Sigma-Model* (NLSM) as the kinetic term of $\phi_c = \phi + i\chi$ (see ref. (48) for more about the NLSM). Since the NLSM target space in general has a nonvanishing curvature, no field redefinition generically exist that could bring the kinetic term to the free (canonical) form with its Kähler potential $K_{\text{free}} = \overline{\Phi}\Phi$.

Let's now consider the full action (156) under the slow-roll condition, ie. when the contribution of the kinetic term is negligible. Then eq. (156) takes the truly chiral form

$$S_{\text{ch.}} = \int d^4x d^2\theta \mathcal{E} \left[X(\Phi)\mathcal{R} + W(\Phi) \right] + \text{H.c.} \tag{164}$$

When choosing X as the independent chiral superfield, $S_{\text{ch.}}$ can be rewritten to the form

$$S_{\text{ch.}} = \int d^4x d^2\theta \mathcal{E} \left[X\mathcal{R} - \mathcal{Z}(X) \right] + \text{H.c.} \tag{165}$$

where we have introduced the notation

$$\mathcal{Z}(X) = -W(\Phi(X)) \tag{166}$$

In its turn, the action (165) is equivalent to the chiral $F(\mathcal{R})$ supergravity action (34), whose function F is related to the function $\mathcal{Z}$ via Legendre transformation (Sec. 6)

$$\mathcal{Z} = X\mathcal{R} - F, \quad F'(\mathcal{R}) = X \quad \text{and} \quad \mathcal{Z}'(X) = \mathcal{R} \tag{167}$$

It implies the equivalence between the reduced action (164) and the corresponding $F(\mathcal{R})$ supergravity whose F-finction obeys eq. (167).

Next, let us consider the special case of eq. (164) when the superpotential is given by

$$W(\Phi) = \frac{1}{2}m\Phi^2 + \frac{1}{6}\lambda\Phi^3 \tag{168}$$

with the real coupling constants $m > 0$ and $\lambda > 0$. The model (168) is known as the *Wess-Zumino* (WZ) model in 4D, $N = 1$ rigid supersymmetry. It has the most general renormalizable scalar superpotential in the absence of supergravity. In terms of the field components, it gives rise to the Higgs-like scalar potential.

For simplicity, let us take a cubic superpotential,

$$W_3(\Phi) = \frac{1}{6}\lambda\Phi^3 \tag{169}$$

or just assume that this term dominates in the superpotential (168), and choose the $X(\Phi)$-function in eq. (164) in the form

$$X(\Phi) = -\xi\Phi^2 \tag{170}$$

with a large positive coefficient ξ, $\xi > 0$ and $\xi \gg 1$, in accordance with eqs. (154) and (155).
Let us also simplify the F-function of eq. (119) by keeping only the most relevant cubic term,

$$F_3(\mathcal{R}) = -\frac{1}{6}f_3\mathcal{R}^3 \tag{171}$$

It is straightforward to calculate the $\mathcal{Z}$-function for the F-function (171) by using eq. (167). We find

$$-X = \frac{1}{2}f_3\mathcal{R}^2 \quad \text{and} \quad \mathcal{Z}'(X) = \sqrt{\frac{-2X}{f_3}} \tag{172}$$

Integrating the last equation with respect to X yields

$$\mathcal{Z}(X) = -\frac{2}{3}\sqrt{\frac{2}{f_3}}(-X)^{3/2} = -\frac{2\sqrt{2}}{3}\frac{\xi^{3/2}}{f_3^{1/2}}\Phi^3 \tag{173}$$

where we have used eq. (170). In accordance to eq. (166), the $F(\mathcal{R})$-supergravity $\mathcal{Z}$-potential (173) implies the superpotential

$$W_{\text{KS}}(\Phi) = \frac{2\sqrt{2}}{3}\frac{\xi^{3/2}}{f_3^{1/2}}\Phi^3 \tag{174}$$

It coincides with the superpotential (169) of the WZ-model, provided that we identify the couplings as

$$f_3 = \frac{32\xi^3}{\lambda^2} \tag{175}$$

We conclude that the original nonminimally coupled matter-supergravity theory (156) in the slow-roll approximation with the superpotential (169) is classically equivalent to the $F(\mathcal{R})$-supergravity theory with the F-function given by eq. (171) when the couplings are related by eq. (175).

The inflaton mass M in the supersymmertic case, according to eqs. (129) and (175), is given by

$$M_{\text{inf}}^2 = \frac{15\lambda^2}{32\xi^3} \tag{176}$$

This relation is different from eq. (151) valid in the bosonic case. Since the value of M_{inf} is fixed by the COBE normalization (Sec. 4), eq. (176) implies that the value of ξ is much lower in the supersymmetric case, $\xi \approx \mathcal{O}(10^{10/3})$, where we have used eqs. (140) and (175), and have assumed that $\lambda \approx \mathcal{O}(1)$.

The established equivalence begs for a fundamental reason. In the high-curvature (inflationary) regime the R^2-term dominates over the R-term in the Starobinsky action (7), while the coupling constant in front of the R^2-action is dimensionless (Sect. 2). The Higgs inflation is based on the Lagrangian (141) with the relevant scalar potential $V_4 = \frac{1}{4}\lambda\phi_J^4$ (the parameter v is irrelevant for inflation), whose coupling constants ξ and λ are also dimensionless. Therefore, both relevant actions are *conformal*. Inflation breaks the conformal symmetry spontaneously.

The supersymmetric case is similar: the nonminimal action (164) with the X-function (170) and the superpotential (169) also has only dimensionless coupling constants ξ and λ, while the same it true for the $F(\mathcal{R})$-supergravity action with the F-function (171), whose coupling constant f_3 is dimensionless too. Therefore, those actions are both *superconformal*, while inflation spontaneously breaks the superconformal invariance.

A spontaneous breaking of the conformal symmetry necessarily leads to Goldstone particle (or dilaton) associated with spontaneously broken dilatations. So, perhaps, Starobinsky scalaron (inflaton) may be identified with the Goldstone dilaton!

The basic field theory model, describing both inflation *and* the subsequent reheating, reads (see eg., eq. (6) in ref. (73))

$$L/\sqrt{-g} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - V(\phi) + \frac{1}{2}\partial_\mu\chi\partial^\mu\chi - \frac{1}{2}m_\chi^2\chi^2 + \frac{1}{2}\tilde{\xi}R\chi^2 + \bar{\psi}(i\gamma^\mu\partial_\mu - m_\psi)\psi$$
$$- \frac{1}{2}g^2\phi^2\chi^2 - h(\bar{\psi}\psi)\phi \tag{177}$$

with the inflaton scalar field ϕ interacting with another scalar field χ and a spinor field ψ. The nonminimal supergravity theory (156) with the Wess-Zumino superpotential (168) can be considered as the $N = 1$ locally supersymetric extension of the basic model (177), after rescaling ϕ_c to $(1/\sqrt{2})\phi_c$ and identifying $\tilde{\xi} = -\frac{1}{3}\xi$ because of eq. (155). Therefore, *pre-heating* (ie. the nonperturbative enhancement of particle production due to parametric resonance (73)) is a generic feature of supergravity models.

The axion χ and fermion ψ are both requred by supersymmetry, being in the same chiral supermultiplet with the inflaton ϕ. The scalar interactions are

$$V_{\text{int}}(\phi, \chi) = m\lambda\phi(\phi^2 + \chi^2) + \frac{\lambda^2}{4}(\phi^2 + \chi^2)^2 \tag{178}$$

whereas the Yukawa couplings are given by

$$L_{\text{Yu}} = \frac{1}{2}\lambda\phi(\bar{\psi}\psi) + \frac{1}{2}\lambda\chi(\bar{\psi}i\gamma_5\psi) \tag{179}$$

Supersymmetry implies the unification of couplings since $h = -\frac{1}{2}\lambda$ and $g^2 = \lambda^2$ in terms of the single coupling constant λ. If supersymmetry is unbroken, the masses of ϕ, χ and ψ are all the same. However, inflation already breaks supersymmetry, so the spontaneously broken supersymmetry is appropriate here.

To conclude, inflationary (slow-roll) dynamics in the gravity theory with a nonminimal scalar-curvature coupling can be equivalent to that in the certain $f(R)$ gravity theory. We extended that correspondence to $N = 1$ supergravity. The nonminimal coupling in supergravity is rewritten in terms of the standard ('minimal') $N = 1$ matter-coupled supergravity, by using their manifestly supersymmetric formulations in curved superspace. The equivalence relation between the supergravity theory with the nonminimal scalar-curvature coupling and the $F(\mathcal{R})$ supergravity during slow-roll inflation is established.

The equivalence is expected to hold even after inflation, during initial reheating with harmonic oscillations. In the bosonic case the equivalence holds until the inflaton field value is higher than $\omega \approx M_{\text{Pl}}/\xi \approx 10^{-5}M_{\text{Pl}}$. In the superymmetric case we have the same bound $\omega \approx M_{\text{Pl}}/\xi^{3/2} \approx 10^{-5}M_{\text{Pl}}$.

12. Reheating and quantum particle production

Reheating is a transfer of energy from inflaton to ordinary particles and fields.

The classical solution (neglecting particle production) near the minimum of the inflaton scalar potential reads

$$a(t) \approx a_0 \left(\frac{t}{t_0}\right)^{2/3} \quad \text{and} \quad \varphi(t) \approx \left(\frac{M_{\text{Pl}}}{3M_{\text{inf}}}\right) \frac{\cos[M_{\text{inf}}(t - t_0)]}{t - t_0} \tag{180}$$

A *time-dependent* classical spacetime background leads to *quantum* production of particles with masses $m < \omega = M_{\text{inf}}$ (71). Actually, the amplitude of φ-oscillations decreases much faster (73), namely, as

$$\exp[-\frac{1}{2}(3H + \Gamma)t] \tag{181}$$

via inflaton decay and the universe expansion, as the solution to the inflaton equation

$$\ddot{\varphi} + 3H\dot{\varphi} + (m^2 + \Pi)\varphi = 0 \tag{182}$$

Here Π denotes the polarization operator that effectively describes particle production. Unitarity (optical theorem) requires $\text{Im}(\Pi) = m\Gamma$. The assumption $m \gg H$ has also been used here (73).

The Starobinsky model in Jordan frame,

$$S = \int d^4x \sqrt{-g_J} f_S(R_J) + S_{\text{SM}}(g^{\mu\nu}{}_J, \psi) \tag{183}$$

after the conformal transformation to Einstein frame reads

$$S = S_{\text{scalar−tensor gravity}}(g_{\mu\nu}, \varphi) + S_{\text{SM}}(g^{\mu\nu}e^{-\sigma\varphi}, \psi) \tag{184}$$

so that the inflaton φ couples to all non-conformal terms and fields ψ, due to the universality of gravitational interaction. Therefore, the Starobinsky inflation also has the *universal* mechanism of particle production.

The perturbative decay rates of inflaton into a pair of scalars (s) or into a pair of fermions (f) are given by (20; 21; 74)

$$\Gamma_{\varphi \to ss} = \frac{M_{\text{inf}}^3}{192\pi M_{\text{Pl}}^2} \qquad \text{and} \qquad \Gamma_{\varphi \to ff} = \frac{M_{\text{inf}} M_f^2}{48\pi M_{\text{Pl}}^2} \tag{185}$$

The perturbative decay rate of inflaton into a pair of *gravitini* is (75)

$$\Gamma_{\varphi \to 2\psi_{3/2}} = \frac{\left|G_{,\varphi}\right|^2}{288\pi} \frac{M_{\text{inf}}^5}{m_{3/2}^2 M_{\text{Pl}}^2} \tag{186}$$

There is *no* parametric resonance enhancement here because the produced particles rapidly scatter. The energy transfers by the time $t_{\text{reh}} \geq \left(\sum_{s,f} \Gamma_{s,f}\right)^{-1}$. One finds the *reheating temperature* (34; 76)

$$T_{\text{reh}} \propto \sqrt{\frac{M_{\text{Pl}}\Gamma}{(\#d.o.f.)^{1/2}}} \approx 10^9 \text{ GeV} \tag{187}$$

that gives the maximal temperature of the primordial plasma.

In the context of supergravity coupled to the supersymmetric matter (like MSSM) gravitino can be either LSP (= the lightest sparticle) or NLSP (= not LSP). In the LSP case (that usually happens with gauge mediation of supersymmetry breaking and $m_{3/2} \ll 10^2$ GeV) gravitino is stable due to the R-parity conservation. If gravitino is NLSP, then it is unstable (it usually happens with gravity- or anomaly- mediation of supersymmetry breaking, and $m_{3/2} \gg 10^2$ GeV). Unstable gravitino can decay into LSP. See ref. (77) for a review of mediation of supersymmetry breaking from the hidden sector to the visible sector.

Stable gravitino may be the dominant part of *Cold Dark Matter* (CDM) (78). There exist severe Big Bang Nucleosynthesis (BBN) [4] constraints on the overproduction of ^3He in that case, which give rise to the upper bound on the reheating temperature of thermally produced gravitini, $T_{\text{reh}} < 10^{5 \div 6}$ GeV (69; 80). The reheating temperature (187) is unrelated to that bound because it corresponds to the much earlier time in the history of the Universe.

When gravitino is NLSP of mass $m_{3/2} \gg 10^2$ GeV, the BBN constraints are drastically relaxed because the gravitino lifetime becomes much shorter than the BBN time (about 1 sec) (69;

---

[4] See ref. (79) for a review of BBN.

80). In that case the most likely CDM candidate is MSSM neutralino, while the reheating temperature may be as high as 10^{10} GeV (80).

An overproduction of gravitini from inflaton decay and scattering processes should be avoided in order to not overclose the Universe. The cosmological constraints on gravitino abundances were formulated in ref. (81). Those constraints are very model-dependent.

The rate of decay changes with time, along with the decreasing amplitude of inflaton oscillations. It stops when the decay rate becomes smaller than the production rate. Then the particle production accelerates (called *pre-heating*, or true BB!) due to the parametric resonance enhancement (73). The reheating rapidly transfers most of energy to radiation, and leads to a radiation-dominated universe with $a \propto t^{1/2}$.

In the matter-coupled $F(\mathcal{R})$ supergravity with the action

$$S = \left[\int d^4x d^2\theta \, \mathcal{E}F(\mathcal{R}) + \text{H.c.}\right] + S_{\text{SSM}}(E, \Psi) \tag{188}$$

after the super-Weyl transformation, $\mathcal{E} \to \mathcal{E}e^{3\Phi}$, we get

$$S = S_{\text{scalar-tensor supergravity}}(E, \Phi) + S_{\text{SSM}}(e^{\Phi+\overline{\Phi}}E, \Psi) \tag{189}$$

so that the superscalaron Φ is *universally* coupled to the SSM matter superfields Ψ.

13. Conclusion

• A manifestly $4D$, $N = 1$ supersymmetric *extension* of $f(R)$ gravity exist, it is *chiral* and is parametrized by a holomorphic function. An $F(R)$ supergravity is classically *equivalent* to the standard theory of a chiral scalar superfield (with certain Kähler potential and superpotential) miminally coupled to the $N = 1$ Poincaré supergravity in four spacetime dimensions (with nontrivial G and K).

• The Starobinsky model of chaotic inflation can be embedded into $F(\mathcal{R})$ supergravity, thus providing the new viable realization of chaotic inflation in supergravity, and the simple solution to the η-problem, by using supergravity only!

The dynamical chiral superfield in $F(\mathcal{R})$ supergravity may be identified with the dilaton-axion chiral superfield in quantum 4D Superstring Theory, when demanding the $SL(2, \mathbf{Z})$ symmetry of the effective action. The $R^2A(R)$ terms may appear in the bosonic gravitational effective action after superstring compactification. The problem is to get the anomalously *large* coefficient in front of the $\mathcal{R}^3$-term in the effective $F(\mathcal{R})$ supergravity theory, that would be consistent with string dynamics.

Supersymmetry in $F(R)$ supergravity is already broken by inflation. The anomaly- or gravitationally-mediated supersymmetry breaking may serve as the important element for the new particle phenomenology (beyond the Standard Model) based on the matter-coupled $F(R)$ supergravity theory.

14. Outlook: CP-violation, baryonic asymmetry, lepto- and baryo-genesis

The observed part of our Universe is $C-$ and $CP-$asymmetric, and it has no antimatter. Inflation naturally implies a *dynamical* origin of the baryonic matter predominance due to a nonconserved baryon number. The main conditions for the dynamical generation of the cosmological baryon asymmetry in the early universe were formulated by A.D. Sakharov in 1967 (62):

1. nonconservation of baryons (*cf.* SUSY, GUT, EW theory),

2. $C-$ and $CP-$symmetry breaking (confirmed experimentally),

3. deviation from thermal equilibrium in initial hot universe.

There exist many scenarios of baryogenesis (see ref. (63) for a review), all designed to explain the observed asymmetry (BBN,CMB):

$$\beta = \frac{n_B - n_{\overline{B}}}{n_\gamma} = 6 \cdot 10^{-10} \tag{190}$$

Here n_B stans for the concentration of baryons, $n_{\overline{B}}$ for the concentration of anti-baryons, and n_γ for the concentration of photons.

Perhaps, the most popular scenario is the nonthermal *baryo-through-lepto-genesis* (64; 65), ie. a creation of lepton asymmetry by L-nonconserving decays of a heavy ($m \approx 10^{10}$ GeV) Majorana neutrino, and a subsequent transformation of the lepton asymmetry into baryonic asymmetry by CP-symmetric, B-nonconserving and (B-L)-conserving electro-weak processes.

The thermal leptogenesis requires high reheating temperature, $T_{\text{reh}} \geq 10^9$ GeV (82), which is consistent with our eq. (187).

The matter-coupled $F(R)$ supergravity theory may contribute towards the *origin* and the *mechanism* of CP-violation and baryon asymmetry, because

  • *complex* coefficients of $F(R)$-function and the complex nature of the $F(R)$ supergravity are the simple *source* of explicit CP-violation and complex Yukawa couplings;

  • the nonthermal leptogenesis is possible via decay of heavy sterile neutrinos (FY-mechanism) *universally produced* by (super)scalaron decays, or via neutrino oscillations in early universe (83);

  • the existence of the natural Cold Dark Matter candidates (*gravitino*, *axion*, *inflatino*) in $F(R)$ supergravity;

  • $F(R)$-supergravity can naturally support *hybrid* (or two-field) inflationary models because it already has a pseudo-scalar superpartner (axion) of inflaton. As is well known, *non-Gaussianity* is a measure of inflaton *interactions* (determined by its 3-point functions and higher). The non-Gaussianity parameter f_{NL} is defined in terms of the (gauge-invariant) comoving curvature perturbations as

$$\hat{R} = \hat{R}_{\text{gr}} + \frac{3}{5} f_{\text{NL}} \hat{R}_{\text{gr}}^2 \tag{191}$$

The non-Gaussianity was not observed yet, though it is expected, while all single-field inflationary models predict $f_{\text{NL}} \approx 0.02$.

15. Acknowledgements

16. References

[1] S.J. Gates, Jr., and S.V. Ketov, Phys. Lett. B674 (2009) 59

[2] S.V. Ketov, Class. and Quantum Grav. 26 (2009) 135006

[3] S.J. Gates, Jr., S.V. Ketov and N. Yunes, Phys. Rev. D80 (2009) 065003

[4] S.V. Ketov, AIP Conf. Proc. 1241 (2010) 613

[5] S.V. Ketov, Phys. Lett. B692 (2010) 272

[6] S. Kaneda, S.V. Ketov and N. Watanabe, Mod. Phys. Lett. A25 (2010) 2753

[7] S. Kaneda, S.V. Ketov and N. Watanabe, Class. and Quantum Grav. 27 (2010) 145016

[8] S.V. Ketov and N. Watanabe, JCAP 1103 (2011) 011

[9] S.V. Ketov and A.A. Starobinsky, Phys. Rev. D83 (2011) 063512

[10] S.V. Ketov and A.A. Starobinsky, *Inflation and nonminimal scalar-curvature coupling in gravity and supergravity*, arXiv:1107.nnnn [hep-th]

[11] A.H. Guth, Phys.Rev. D23 (1981) 347

[12] A.D. Linde, *Particle Physics and Inflationary Cosmology*, Harwood, Chur, Switzerland, 1990

[13] A.R. Liddle and D.H. Lyth, *Cosmological Inflation and Large-scale Structure*, Cambridge University Press, Cambridge, 2000

[14] E. Komatsu et al., Astrophys. J. Suppl. 192 (2011) 18; http://map.gsfc.nasa.gov

[15] D.N. Spergel et al., Astrophys. J. Suppl. 170 (2007) 377

[16] L.D. Landau and E.M. Lifshitz, *The Classical Theory of Fields*, Pergamon Press, Oxford, 2002

[17] V.F. Mukhanov, *Physical Foundations of Cosmology*, Cambridge University Press, 2005

[18] A.R. Liddle and D.H. Lyth, *The Primordial Density Perturbation: Cosmology, Inflation and the Origin of Structure*, Cambridge University Press, 2009

[19] D.S. Gorbunov, *Introduction to the Theory of the Early Universe: Cosmological Perturbations and Inflationary Theory*, World Scientific, 2010

[20] A.A. Starobinsky, Phys. Lett. B91 (1980) 99

[21] A.A. Starobinsky, *Nonsingular model of the Universe with the quantum-gravitational de Sitter stage and its observational consequences*, in the Proceedings of the 2nd Intern. Seminar "Quantum Theory of Gravity" (Moscow, 13–15 October, 1981); INR Press, Moscow 1982, p. 58 (reprinted in "Quantum Gravity", M.A. Markov and P.C. West Eds., Plemum Publ. Co., New York, 1984, p. 103)

[22] R.P. Woodard, Lect. Notes Phys. 720 (2007) 403

[23] A.D. Linde, Phys. Lett. B 129 (1983) 177

[24] D. Wands, Class. and Quantum Grav. 11 (1994) 269

[25] T.P. Sotiriou and V. Varaoni, Rev. Mod. Phys. 82 (2010) 451

[26] A. De Felice and S. Tsujikawa, Living Rev. Rel. 13 (2010) 3

[27] V. Müller, H.-J. Schmidt and A.A. Starobinsky, Phys. Lett. B202 (1988) 198

[28] B. Whitt, Phys. Lett. B145 (1984) 176

[29] J.D. Barrow and S. Cotsakis, Phys. Lett. B214 (1988) 515

[30] K.-I. Maeda, Phys. Rev. D39 (1989) 3159

[31] S. Dodelson, *Modern Cosmology*, Elsevier, 2003

[32] S. Weinberg, *Cosmology*, Oxford Univ. Press, 2008

[33] V.F. Mukhanov and G.V. Chibisov, JETP Lett. 33 (1981) 532

[34] A.A. Starobinsky, Sov. Astron. Lett. 9 (1983) 302

[35] G.F. Smoot, Astrophys. J. 396 (1992) L1;
      http://en.wikipedia.org/wiki/Cosmic_Background_Exprorer
[36] A. Linde, M. Noorbala and A. Westphal, JCAP 1103 (2011) 013
[37] E. Hand, Nature 458 (2009) 820
[38] S.J. Gates, Jr., M.T. Grisaru, M. Roček and W. Siegel, *Superspace or 1001 Lessons in Supersymmetry*, Benjamin-Cummings Publ. Company, 1983
[39] J. Wess and J. Bagger, *Supersymmetry and Supergravity*, Princeton Univ. Press, 1992
[40] I.L. Buchbinder and S. M. Kuzenko, *Ideas and Methods of Supersymmetry and Supergravity*, IOP Publ., 1998
[41] S.J. Gates, Jr., Phys. Lett. B365 (1996) 132
[42] S.J. Gates, Jr., Nucl. Phys. B485 (1997) 145
[43] S. Cecotti, Phys. Lett. B190 (1987) 86
[44] P.S. Howe and R.W. Tucker, Phys. Lett. B80 (1978) 138
[45] E. Cremmer, B. Julia, J. Scherk, S. Ferrara, L. Girardello and P. van Nieuwenhuizen, Nucl. Phys. B147 (1979) 105
[46] M. Yamaguchi, Class. and Quantum Grav. 28 (2011) 103001
[47] H. Murayama, H. Suzuki, T. Yanagida and J. Yokoyama, Phys. Rev. D50 (1994) 2356
[48] S.V. Ketov, *Quantum Non-linear Sigma-models*, Springer-Verlag, 2000
[49] M. Kawasaki, M. Yamaguchi and T. Yanagida, Phys. Rev. Lett. 85 (2000) 3572
[50] P.B. Binetruy and G. Dvali, Phys. Lett. B388 (1996) 241
[51] P. Berglund and G. Ren, *Multi-field inflation from string theory*, arXiv:0912.1397 [hep-th]
[52] P. Berglund and G. Ren, *Non-Gaussianity in string cosmology: a case study*, arXiv:1010.3261 [hep-th]
[53] M. R. Douglas and S. Kachru, Rev. Mod. Phys. 79 (2007) 733
[54] R. Kallosh, Lecture Notes Phys. 738 (2008) 119
[55] E. Cremmer, S. Ferrara, C. Kounnas and D. V. Nanopoulos, Phys. Lett. B133 (1983) 61
[56] S. Kachru, R. Kallosh, A. Linde and S. P. Trivedi, Phys. Rev. D68 (2003) 046005
[57] K.-I. Maeda, Phys. Rev. D37 (1988) 858
[58] A.L. Berkin and K.-I. Maeda, Phys. Lett. B245 (1990) 348
[59] M. Abramowitz and I.A. Stegun, Eds., *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, New York, Dover, 1972, p. 17
[60] L.A. Kofman, A.D. Linde and A.A. Starobinsky, Phys. Lett. B157 (1985) 361
[61] A.A. Starobinsky, JETP Lett. 34 (1981) 438
[62] A.D. Sakharov, Pis'ma ZhETF 5 (1967) 32
[63] M. Dine and A. Kusenko, Rev. Mod. Phys. 76 (2004) 1
[64] W. Buchmüller, R.D. Peccei and T. Yanagida, Ann. Rev. Nucl. Part. Sci. 55 (2005) 311
[65] M. Fukugita and T. Yanagida, Phys. Lett. B174 (1986) 45
[66] W. Buchmüller, K. Schmitz and G. Vertongen, Phys. Lett. B693 (2010) 421
[67] F.L. Bezrukov and M. Shaposhnikov, Phys. Lett. B659 (2008) 703
[68] F. Bezrukov, AIP Conf. Proc. 1241 (2010) 511
[69] M. Kawasaki, K. Kohri and K. Moroi, Phys. Rev. D71 (2005) 083502
[70] F. Bezrukov, A. Magnin, M. Shaposhnikov and S. Sibiryakov, JHEP 1101 (2011) 016
[71] N.D. Birrell and P.C.W Davies, *Quantum Fields in Curved Space*, Cambridge Univ. Press, 1982
[72] M.B. Einhorn and D.R.T. Jones, JHEP 1003 (2010) 026
[73] L. Kofman, A. Linde and A.A. Starobinsky, Phys. Rev. D56 (1997) 3258
[74] A. Vilenkin, Phys. Rev. D32 (1985) 2511

[75] M. Endo, K. Hamaguchi and F. Takahashi, Phys. Rev. Lett. 96 (2006) 211301

[76] D.S. Gorbunov and A.G. Panin, *Scalaron the mighty: is a quick mare in time everywhere?*, arXiv:1009.2448 [hep-ph]

[77] G.F. Giudice and R. Rattazzi, Phys. Rep. 322 (1999) 419

[78] M. Bolz, A. Brandenburg and W. Buchmüller, Nucl. Phys. B606 (2001) 518

[79] F. Iocco, G. Mangano, G. Miele, O. Pisanti and P.D. Serpico, Phys. Rep. 472 (2009) 1

[80] M. Kawasaki, K. Kohri, T. Moroi and A. Yotsuyanagi, Phys. Rev. D78 (2008) 065011

[81] M. Endo, F. Takahashi and T.T. Yanagida, Phys. Rev. D76 (2007) 083509

[82] W. Buchmüller, P. Di Bari and M. Plumacher, Annals of Phys. 315 (2005) 305

[83] E.K. Akhmedov, V.A. Rubakov and A.Y. Smirnov, Phys. Rev. Lett. 81 (1998) 1359.

Supersymmetric Dark Matter

Csaba Balázs and Daniel Carter
Centre of Excellence for Particle Physics at the Terascale and
Monash Centre for Astrophysics
School of Physics, Monash University
Australia

1. Introduction

The standard model of particle physics excellently describes all known fundamental particles and forces (with the exception of gravity). Over the last several decades, this theory has become established through a preponderance of experimental data, culminating in the detection of the top quark at the Tevatron collider. However, while there has never been a direct observation of a particle that does not fit in the standard model, there are compelling reasons to believe that such particles exist, and so the standard model is not a complete description of Nature.

The most noteworthy piece of evidence in this case comes from astrophysics: the enigmatic dark matter. Dark matter was postulated to explain why stars in galaxies appeared to orbit faster than the gravitational attraction of the luminous mass would imply. This hypothesis, that there is more matter in the universe than is luminous, has been confirmed by observations in galactic structure formation, gravitational lensing, and most importantly in the cosmic microwave background radiation. Relics from the early universe, small deviations from uniformity in the cosmic microwave background radiation provide details on the matter content of the universe, and in particular it places stringent upper bounds on the amount of baryonic matter, ruling out atomic matter as an explanation for dark matter. Ultimately, it has been determined that any potential particle of dark matter must satisfy several requirements. It must be massive, apart from gravitation it may only interact very weakly, it must be stable on a cosmic time scale, and it must be non-relativistic (having average temperature much below its mass scale). This immediately excludes any particle in the standard model.

Other possible shortcomings with the standard model are also emerging. For example, in certain precision experiments, such as the measurement of the anomalous magnetic moment of the muon, discrepancies seem to be appearing between the experimentally measured value and the value predicted by the standard model. These results may indicate the presence of additional contributions from currently unknown particles.

We are therefore led to consider new theories that extend the standard one. Any such theory should explain these new phenomena, but must also reproduce the myriad successes of the standard model. One of the most promising ideas is the extension of the standard model by the introduction of supersymmetry. Supersymmetry is the unification of fermions and bosons, in other words the unification of matter and force. Under supersymmetry, all fermions and bosons are considered as components of a superfield, and hence for every fermionic field in a

supersymmetric theory there will be a corresponding bosonic field (and vice versa) with the same mass and quantum numbers. If we impose supersymmetry on the standard model, we therefore necessarily introduce a number of new particles: a new spin zero field for both spin degrees of freedom for each quark and lepton, a new spin-half field for each of the force carrier bosons, and additional spin-half fields for the complex degrees of freedoms of the Higgs bosons. This gives us the Minimal Supersymmetric Standard Model (MSSM). From the new electrically neutral fermionic fields we construct the neutralinos, the lightest of which satisfies all the basic requirements for being a dark matter particle. As well as this, supersymmetry neutralises the quadratic divergences in the quantum corrections to the Higgs mass, solving the so-called hierarchy problem.

Of course, supersymmetry is not preserved in the universe today, or we would certainly have detected the presence of at least some super-particles. We know supersymmetry must be broken. However the precise mechanism can so far only be speculated. One particularly interesting idea is supergravity, in which supersymmetry is elevated to a local symmetry and as a result is broken via gravity. Our investigations concern the Next to Minimal Supersymmetric Standard Model (NMSSM) in which the MSSM is augmented by a gauge singlet superfield subject to discrete symmetries, which alleviates naturalness issues of the Higgs mass parameter. On this scenario we impose supergravity constraints, creating the NmSuGra model.

Unsurprisingly though, testing a new theory for viability can pose a considerable challenge. Not only is it highly complex to calculate predictions for a given model, but invariably there will be degrees of freedom in the model which are not constrained by theoretical considerations, which must therefore become free parameters. Varying these creates a multiplicity of possibilities that can easily become unmanageable, and so sophisticated statistical techniques are required to keep track of supersymmetric predictions. We adopt Bayesian probability theory as the appropriate general formalism for examining the plausibility of a model, and our results employ sophisticated numerical algorithms to explore the parameter space of the theory. For each set of parameters, predictions are calculated for a comprehensive selection of pertinent observables, including dark matter abundance, collider limits and precision measurements, which are used to assign a likelihood value.

While there have not so far been any direct observations of dark matter or of supersymmetry, there are good prospects for such events to occur in the near future. The CERN Large Hadron Collider (LHC) will have the capacity to create supersymmetric particles if they are light enough, and may produce detectable signals in the form of missing energy from escaped neutralinos. Furthermore, experiments exist with the aim of detecting cosmic dark matter directly through nuclear scattering events in underground vats. We find that the chances of detection for NmSuGra (and a broad class of similar theories) at these experiments are quite favourable.

2. Dark matter

The existence of dark matter is one of the most significant enduring mysteries of modern particle astrophysics, every bit as enigmatic as its name suggests. Dark matter is the generic name for an unknown substance which is possibly made of more than one constituents. It has never been directly observed, but it is now believed to constitute the bulk of matter existing in the Universe today.

Experimental evidence for the existence of dark matter is overwhelming. Historically, dark matter was proposed in order to explain a number of anomalous astrophysical observations in which the effects of gravity could not be accounted for by the measured mass of luminous matter. Initial evidence for dark matter came from the curious observation that outer parts of galaxies rotate faster than they should. In the seminal observations of Vera Rubin, the orbital velocities of stars were measured as a function of distance from the center of the Andromeda Galaxy (Rubin & Ford, 1970). Rubin and collaborators determined that the velocities did not vary greatly with distance, whereas the centralised distribution of visible mass would be consistent with velocities decreasing rapidly as radial distance increases. This discrepancy could be resolved by the presence of a galactic halo of some non-visible, gravitating matter, that has subsequently come to be known as dark matter.

Numerous other observations, originating from vastly different length scales, support the dark matter hypothesis. Light bending around the enormous gravitational potential of Galaxy super-clusters, an effect coined as gravitational lensing, indicates significantly more mass within these clusters that meets the eye (Kneib et al., 1996). Simulations of Galaxy and larger structure formation have determined that the clustering of galaxies observed in the Universe today could not occur under the gravitational effects of the matter that can be seen (Abazajian et al., 2009). An interesting special case, the Bullet Cluster, is comprised of two galactic clusters undergoing a violent, cosmic collision. Individual galaxies and intergalactic gas, consisting of charged cosmic rays, are visible (the latter mainly due to its X-ray emissions). Via gravitational lensing it was observed that the concentration of mass was separated from the visible matter in the Bullet Cluster. In fact, the bulk of mass was carried by an invisible, non-luminous agent (Markevitch et al., 2004). A simple explanation of these observations is that visible matter, which interacts more strongly than dark matter, was left behind closer to the core of the collision while dark matter halos surrounding them, interacting weaker, rushed forward and the two became separated from each other (Clowe et al., 2006).

The most precise value for the amount of dark matter in the Universe comes from measurements of anisotropies in the cosmic microwave background, the CMB. In the early Universe, after quarks bound into atomic nuclei, matter existed in an opaque plasma state. As the Universe cooled and neutral atoms were formed, the thermal radiation produced by this plasma was able to propagate freely. Today we can measure the remnant of this radiation as a highly uniform microwave signal observable from all directions of the sky. The small deviations from uniformity, density fluctuations in the early plasma, are the cosmic seeds of structure formation. The amount of the non-uniformity depends on the amount of matter in the early Universe. This is because more matter gravitates stronger thus, collapsing under the effect of stronger gravity, more matter creates higher density fluctuations. Thus the power spectrum of the CMB can be used to deduce the matter composition of the Universe (Komatsu et al., 2011).

The CMB spectrum also gives us the density of ordinary atomic matter built of baryons, which are bound states of quarks. From the CMB we can deduce that only about twenty percent of the total matter density is baryonic. This is confirmed, with excellent precision, by the observation of the abundances of the lightest elements in the Universe (Amsler et al., 2008). The fact that baryonic energy density is only a fraction of the total matter density excludes the simplest possibility that dark matter is composed of regular atomic matter.

But then what is dark matter made of? To address this question we have to ask: What do we know about the properties of dark matter so far? From astrophysical measurements

we deduced that dark matter is non-baryonic. It must also be stable against decay, at least on the timescale of the lifetime of the Universe, otherwise it could not change gravitational observations today. Structure formation further indicates that dark matter is cold, which means that the speed of the thermal motion of dark matter constituents at the present average temperature of the Universe is much lower than the speed of light. Finally, from observations such as the Bullet Cluster we know that dark matter must be very weakly interacting. It may only interact via gravity or a similarly feeble force.

The above properties exclude the possibility that the bulk of dark matter is made of known stable particles such as protons, electrons or neutrinos. Thus, there are numerous competing theories and speculations regarding what dark matter might be made of. It appears that one of the simplest avenues toward solving the dark matter puzzle is to postulate the existence of an undiscovered, exotic particle. The density evolution of such dark matter particles over the eons can be traced by thermodynamical calculations (Kolb & Turner, 1990). These calculations show that dark matter density decreased over the history of the Universe due to expansion of space and annihilation between dark matter particles and anti-particles. At some temperature in the early Universe, known as the freeze-out temperature, the density of dark matter decreased to the point that annihilation became negligible to the further evolution of the density, which is now dominated by expansion. What remains today is called the relic abundance, or relic density, of dark matter. The present dark matter abundance depends strongly on the specific properties of the dark matter particle, and is thus essential for discriminating between different theories of dark matter.

Another important question is: How can we observe dark matter particles? Three main avenues were suggested so far: indirect, direct and collider detection. Indirect detection of dark matter utilizes the fact that annihilation to standard particles continues, particularly in localised high-density regions that re-collapsed under the effects of gravity. As today dark matter particles are cold, their non-relativistic thermal kinetic energy is negligible compared to their mass. Thus, the annihilation energy is dominated by the mass of the dark matter particle. Hence it is anticipated that dark matter annihilation will produce standard particle pairs at energies equaling the mass of the dark matter particle. The hope is to detect such a dramatic, peaked signal on the background of the relatively smooth cosmic ray flux of standard astrophysical origin.

Direct dark matter detection experiments aim to find traces of dark matter in ground based detectors. In these highly shielded detectors dark matter particles may bounce into atomic nuclei, which is typically a pure noble gas or a crystalline metal. The boosted nuclei in turn dissipate their energy in the form of heat, light or ionization. Extremely sensitive phonon, scintillation and ionization detectors then can observe this signal of dark matter-nucleon collision.

The ultimate method of discovering dark matter particles is producing them in the laboratory. If dark matter particles are heavy, relativity dictates that this requires colliding standard particles at large energies. If dark matter interactions are very weak, then due to the probabilistic nature of quantum physics, the process also necessitates many of these collisions. The highest energy and luminosity particle collider presently is the Large Hadron Collider at CERN. So all eyes fall on the LHC when it comes to making dark matter on Earth. Eventually, answering the question of what dark matter actually is requires the synthesis of particle physics with astrophysics, the understanding of phenomena at the largest and the smallest scales in the same time.

3. Supersymmetry

Over the second half of the twentieth century, the standard model of particle physics became established as the theoretical framework describing everything that was known about particle physics. With the exception of gravity the standard model incorporates all the fundamental forces that have ever been observed: the electromagnetic, weak and strong forces. These forces are physical consequences of the gauge symmetries of the standard model. Corresponding to these symmetries are three types of force carriers: the photon, weak bosons and gluons. Matter in the standard model is represented by fermions: six quarks and six leptons which are arranged into three generations. An essential, albeit curious, feature of the standard model is that weak interactions violate chirality: electrically charged weak bosons only mediate forces between left-handed fermions. Due to this the standard model exhibits a chiral symmetry, that is the left- and right-handed fermions transform independently. All these features and properties of the standard model have been repeatedly confirmed in experiment, some with a remarkable precision.

However, in recent years data has been accumulating for which the standard model is insufficient to explain. The existence of dark matter is one of the most compelling evidences. A dark matter particle must be electrically neutral and it cannot be baryonic. It must also be stable and have sufficient mass. This leaves no viable candidate for dark matter in the standard model. There are other experimental indications that the standard model may no longer be sufficient, some of which will be discussed in section 4.

Perhaps even more importantly there are conceptual issues the resolution of which lies outside the scope of the standard model. These include the origin of fundamental particle properties such as mass, spin and their various charges such as electromagnetic, weak and strong charges. A further puzzle is the number and variety of elementary particles: six types of quarks and leptons, and three types of force carriers.

The most immediate conceptual problem of the standard model is the underlying dynamics of electroweak symmetry breaking. Electroweak symmetry breaking generates the masses of all elementary particles and the necessary agent of this is the Higgs boson. The Higgs boson, so far a hypothetical particle, is one of a kind – the only spin zero fundamental particle. Such scalar particles spell trouble in the standard model. The standard gauge and chiral structure protects the masses of matter particles and the force carriers. In any quantum field theory, such as the standard model, particle masses receive higher order corrections from vacuum fluctuations. These corrections can grow essentially unbounded unless there is a mechanism to limit them. There is no such mechanism for the Higgs boson mass in place in the standard model. Unless something protects the Higgs mass, electroweak symmetry breaking collapses in the standard model. This is called the hierarchy problem because the unbounded growth of the Higgs mass is thought to be due to the large difference between the weak and Planck energy scales.

It turns out that it is possible to utilize the existing structure of the standard model to extend the protection to the Higgs mass. There would be no hierarchy problem if either the existing gauge or chiral symmetries of the standard model could be used to protect the Higgs mass. This is only wishful thinking unless the Higgs mass is directly connected to one of these symmetries. The introduction of a new overarching symmetry, supersymmetry, does just this (Baer & Tata, 2006). Supersymmetry demands that the Higgs boson is partnered with a chiral fermion, called the Higgsino. The Higgs and Higgsino masses have to be the same due to supersymmetry. But since the Higgsino is a chiral fermion its mass is protected by the

standard chiral symmetry. Thus the introduction of supersymmetry extends the protection to the Higgs mass.

Supersymmetry is attractive for various other reasons. It is the only possible extension of the Poincaré symmetry of space-time that is consistent with quantum field theory. Formally, we can define an operator Q that acts on a fermionic state and transforms it to a bosonic state, and vice versa. It follows that Q must therefore be a spin-half operator, obeying anti-commutation relations. It can be shown that the generators of the Poincaré symmetry accompanied by the generators of the supersymmetry transformations extend the Poincaré group to a super-structure. Amongst other things this implies that a specific combination of supersymmetric transformations is simply a space-time shift. Such extensions of the Poincaré group are in general forbidden by the Coleman-Mandula theorem, however this theorem does not cover non-commuting operators like Q (Coleman & Mandula, 1967). According to the more general Haag-Lopuszanski-Sohnius theorem, supersymmetry is the unique possible extension of the Poincaré algebra, which is quite intriguing (Haag et al., 1975).

Supersymmetry posits that there exists a fundamental symmetry between fermionic and bosonic fields, that in fact these are merely components of a more fundamental entity called a superfield. There are two principle kinds of superfields: chiral and vector superfields. The former contain two physical scalar fields and one physical spin-half field, and are used to represent the standard model fermion fields and any scalar field such as the Higgs boson. The latter embody a (massless) vector boson and a spin-half field, and represent the standard model vector fields (before particles acquire masses). To construct a realistic supersymmetric theory, we take all the fields in the standard model and elevate them to an appropriate superfield, thereby introducing a fermionic counterpart for every boson and two scalar counterparts for every fermion. This then forms the particle content of the Minimal Supersymmetric Standard Model (MSSM).

The simplest supersymmetric particle Lagrangian is the early model by Wess & Zumino (1974). The Wess-Zumino model describes one fermionic and two bosonic fields, or in other words a single superfield. The MSSM, the simplest supersymmetric extension of the standard model, was first constructed by Dimopoulos & Georgi (1981), and remains the most popular supersymmetric model today. The success of the MSSM is due to its versatility. The MSSM was proposed to solve the hierarchy problem but it turned out to solve various other problems of the standard model. It, for example, contains one or more viable dark matter particles. The lightest supersymmetric particle is typically long lived, massive, electrically neutral and weakly interacting: an ideal dark matter candidate (Goldberg, 1983; Pagels & Primack, 1982). In the MSSM various particles, such as the gravitino, the lightest nautralino, the lighest sneutrino, or even the axino can be dark matter. Unification of gauge couplings also substantially improves the MSSM compared to the standard model (Amaldi et al., 1991).

In general the super-partners will mix to form distinct mass eigenstates. The mass eigenstates of the super-partners of the electrically neutral standard electroweak bosons, for example, are called neutralinos and their electrically charged counterparts are the charginos. The lightest neutralino in the MSSM is comprised of the neutral components of the two Higgs superfields (Higgsinos) and the counterparts to the weak W^0 and B^0 bosons. Spin zero super-partners are identified by the names of their spin-half counterparts with an 's' prepended, such as 'selectron' for the scalar partners of the electron.

Super-partners are clearly distinguished from their standard counterparts by a quantum number called R-parity. The R-parity of the fields in the MSSM are defined as

$$R = (-1)^{3(B-L)+2S} \tag{1}$$

where B, L and S represent the baryon number, lepton number and spin of the field, respectively. If we assume that R-parity is a conserved quantum number, then the lightest supersymmetric particle becomes stable against decay.

The form of a supersymmetric Lagrangian including interactions between chiral and vector superfields is largely fixed by the gauge and chiral symmetries of the standard model, and the Poincaré symmetry extended to supersymmetry. Matter-Higgs interactions of a supersymmetric theory are specified by the superpotential, which is a function of renormalizable combinations of left-chiral superfields. The superpotential defining the MSSM is:

$$W_{MSSM}|_{\mu=0} = \mu \hat{H}_u \cdot \hat{H}_d + y_t \hat{Q}_L \cdot \hat{H}_u \hat{T}_R^c - y_b \hat{Q}_L \cdot \hat{H}_d \hat{B}_R^c - y_\tau \hat{L}_L \cdot \hat{H}_d \hat{L}_R^c \tag{2}$$

The left-handed superfields $\hat{Q}_L$, $\hat{L}_L$ and $\hat{H}_{u,d}$ contain the left-handed (s)quark, (s)lepton and Higgs fields respectively, and so are weak $SU(2)_L$ doublets, whereas the right-handed $\hat{T}_R$, $\hat{B}_R$, $\hat{L}_R$ superfields are singlets under weak transformations. Generation indices have been suppressed for the matter fields, but in general the Yukawa couplings y_t are 3×3 matrices in generation space.

Unlike the Yukawa couplings that originate from electroweak symmetry breaking the coefficient μ of the Higgs self-interaction term is a supersymmetric parameter. In theory its natural size is not linked to the electroweak scale at all. On the other hand in the MSSM electroweak symmetry breaking will not happen unless the the size of μ is close to the electroweak scale. This issue can be addressed, however, by considering extensions of the minimal superpotential. In the so-called next-to-minimal supersymmetric standard model we introduce a gauge-singlet superfield (Fayet, 1975), augmenting the μ term in the MSSM superpotential with:

$$W_{NMSSM} = W_{MSSM}|_{\mu=0} + \lambda \hat{S} \hat{H}_u \cdot \hat{H}_d + \frac{\kappa}{3} \hat{S}^3, \tag{3}$$

The scalar component S of the singlet superfield acquires a vacuum expectation value $s = \langle S \rangle$ after electroweak symmetry breaking, generating $\mu = \lambda s$, where λ is an order 1 coupling. Since the size of s is naturally electroweak the μ problem of the MSSM is automatically solved. As there are no gauge couplings for $\hat{S}$, the impact on phenomenology compared to the MSSM is small, although the possibility of the singlet field mixing with the scalar fields could weaken couplings to other fields, making it harder to detect.

In a supersymmetric field theory, for every fermionic degree of freedom there must exist one bosonic degree of freedom, and vice versa, with otherwise identical quantum numbers and masses. For example, in a supersymmetric quantum electrodynamics there would exist two complex spin zero counterparts to the electron for its two spin degrees of freedom. If such super-partners existed in our observed world they would have been seen in experiments well before now, so we are lead to the conclusion that supersymmetry, if it is manifested in our Universe, must be broken at some early time.

How supersymmetry is broken can only be speculated at present. A typical assumption is that the, probably spontaneous, breaking of supersymmetry is communicated from a hidden sector to us at a very high energy scale close to the one where the strength of all fundamental forces unify. When supersymmetry breaking happens super-partners acquire masses. If no new physics occurs between the scale of supersymmetry breaking and the electroweak scale then the renormalization group equations of the MSSM allow us to to calculate the evolution of the super-partner masses and couplings between the two scales. In turn this enables us to calculate the super-partner masses and couplings at the energy scale which is probed by experiments.

Typically supersymmetry is broken such that the breaking does not ruin the protection of the Higgs, and other spin zero super-partner, masses (Dimopoulos & Georgi, 1981). This soft breaking also decouples the origin of supersymmetry breaking from its phenomenological consequences. It is parametrized by adding the following explicit symmetry breaking to the supersymmetric Lagrangian:

$$\mathcal{L}_{NMSSM}^{soft} = \mathcal{L}_{MSSM}^{soft}|_{B=0} - M_S^2|\tilde{S}|^2 - (\lambda A_\lambda \tilde{S} H_u \cdot H_d + \frac{\kappa A_\kappa}{3} \tilde{S}^3 + h.c.). \tag{4}$$

Here

$$\mathcal{L}_{MSSM}^{soft}|_{B=0} = \mathcal{L}_{gaugino}^{soft} + \mathcal{L}_{scalar}^{soft} + \mathcal{L}_{tri-linear}^{soft}, \tag{5}$$

contains the mass terms for the twelve gauginos ($i = 1...3$, $a = 1...8$)

$$\mathcal{L}_{gaugino}^{soft} = -\frac{1}{2}(M_1 \bar{\tilde{B}}\tilde{B} + M_2 \bar{\tilde{W}}_i \tilde{W}_i + M_3 \bar{\tilde{G}}_a \tilde{G}_a + h.c), \tag{6}$$

the sfermions and Higgses

$$\mathcal{L}_{scalar}^{soft} = -(\ M_Q^2|\tilde{Q}|^2 + M_{T_R}^2|\tilde{T}_R|^2 + M_{B_R}^2|\tilde{B}_R|^2 + M_L^2|\tilde{L}|^2 + M_{L_R}^2|\tilde{L}_R|^2 +$$

$$M_{H_u}^2|H_u|^2 + M_{H_d}^2|H_d|^2), \tag{7}$$

and the so called soft tri-linear terms

$$\mathcal{L}_{tri-linear}^{soft} = -(y_t A_t \tilde{Q} \cdot H_u \tilde{T}_R^c - y_b A_b \tilde{Q} \cdot H_d \tilde{B}_R^c - y_\tau A_\tau \tilde{L} \cdot H_d \tilde{L}_R^c + h.c.). \tag{8}$$

The most well studied supersymmetry breaking mechanism is known as the minimal supergravity motivated (mSuGra) model, in which supersymmetry breaking is mediated via gravity and the supersymmetry breaking scale is identified by the unification scale (Chamseddine et al., 1982). The idea behind supergravity is simple and elegant: supersymmetry is promoted to a local symmetry. As supersymmetry is an extension of the Poincaré space-time symmetry, elevating supersymmetry to a local symmetry, we automatically incorporate gravity. The free parameters of minimal supergravity, introduced at the unification scale and appearing in the supersymmetry breaking Lagrangian, are as follows:

$$P = \{M_0, M_{1/2}, A_0, \tan\beta, \lambda\}. \tag{9}$$

Here M_0 is the unified scalar mass, $M_{1/2}$ is the unified gaugino mass, A_0 is a unified trilinear coupling, and $\tan \beta$ is the ratio of the vacuum expectation values of the two Higgs fields, an expedient replacement for the Higgs bilinear term B. For the next to minimal equivalent (NmSuGra) we add λ, the coupling constant in equation 3. Together with experimentally measured standard model input values, these parameters completely define the particle mass spectrum, interactions and mixings required to compute Feynmann diagrams, and hence derive predictions for quantities of interest.

4. Supersymmetric predictions

Supersymmetry is a predictive theoretical framework. Once the values of the parameters have been fixed the Lagrangian is unique and we are able to make various types of theoretical predictions. The most important of these predictions are the masses of the super-partners. This set of masses is frequently referred to as the supersymmetric spectrum and computer programs calculating it called spectrum generators. The calculation of the spectrum involves two main steps. The first is following the renormalization evolution of the supersymmetric masses from the high energy scale, which is typically the unification scale, down to the low scale, which is typically the TeV scale probed by modern experiments (Martin & Vaughn, 1994). In the numerical implementation a spectrum generator solves the relevant renormalisation group equations and runs the masses down from the unification scale to the weak scale. In the second step the eigenvalues of the mass matrices have to be found by diagonalizing, leading to the supersymmetric mass eigenstates. In the process of numerically solving the renormalization group equations the couplings of these mass eigenstates are also identified.

Knowledge of the supersymmetric spectrum permits the calculation of decay rates, scattering amplitudes and related probabilities. These, in turn, allow us to calculate various observables the predicted values of which can be compared to experimental measurements. Foremost among these observables is the relic abundance, the average energy density of the lightest supersymmetric particle which typically forms part of dark matter. In supergravity motivated models the lightest supersymmetric particle is the lightest neutralino. The present energy density of these neutralinos depend on the supersymmetric spectrum and couplings. Knowing these the neutralino abundance can be evolved from high temperatures to the freeze-out temperature. This requires the calculation of neutralino self-scattering (annihilation) cross sections into standard particles and neutralino scattering (co-annihilation) cross sections on other supersymmetric particles. The size of these cross sections depend on the scattering energy and thus the average temperature of the early Universe. They enter into Boltzmann's equation, which in turn determines the neutralino relic abundance as the function of temperature. The value of this abundance after freeze-out is a very important constraint on the choice of parameters in supersymmetric models.

Based on the supersymmetric spectrum cross sections of dark matter scattering on atomic nuclei can also be calculated. In supersymmetric models dark matter scattering on nuclei is mostly mediated by Higgs boson exchange. This spin-independent process is augmented by the spin-dependent Z boson and pseudo-scalar exchange. Neutralinos directly couple to Higgses (or pseudo-scalars) which couple to nuclei via virtual heavy-quark loops. Neutralino-nucleon scattering can also be mediated by scalar partners of quarks. These theoretical cross sections can be compared to measurements of experiments that attempt to

observe dark matter directly as it passes through the Earth. Such detectors are placed deep underground in order to shield them from ordinary particles, thus minimising background. Even this way charged particles created by high energy cosmic neutrinos can reach these detectors so clever event vetoing has to be applied. Since the dark matter flux is expected to vary throughout the year, as the earth orbits the sun, an annual modulation in the signal could help identifying signals of dark matter.

Another avenue to learn about dark matter particles is indirect detection. Dark matter particles can annihilate with their own anti-particles to produce charged cosmic rays or gamma rays in the Galaxy. Such cosmic rays are expected to arrive to us from regions with high dark matter number density such as the center of our Galaxy. These cosmic rays can be detected by satellites specifically designed to measure them. Recently the Fermi-LAT and PAMELA satellites enabled us to very precisely measure the cosmic electron, positron, and gamma ray fluxes as the function of their energy. These measurements are important constraints on supersymmetric models since dark matter annihilations produce electron-positron and photon pairs at energies equaling the mass of a dark matter particle. After the cosmic rays are produced and they propagate through the inter-galactic medium they lose energy. This energy loss can be followed through solving a diffusion equation with reasonable precision. Thus the shape of the observed cosmic ray spectra may carry information to us about the mass of the dark matter particle. The amplitude of a potential dark matter annihilation signal in cosmic rays depends on the number density of the dark matter particles and their annihilation cross-section into the given cosmic ray species. While the dark matter density is relatively well predicted in the vicinity of our solar system it is quite uncertain toward the center of the Galaxy. Thus, the size of an indirect detection signal suffers from various astrophysical uncertainties. Nevertheless, they already provide useful constraints on supersymmetric models.

High energy colliders have the best chance to establish the existence of supersymmetry. Because they are highly controlled experiments they can provide us with the most comprehensive information about a given supersymmetric model. The center of mass energies of current generation colliders are sufficient to reach a significant portion of the parameter space of various supersymmetric models. The LHC, as the highest energy and luminosity collider, leads the way in the search for supersymmetry. Super-particles most likely to be produced at the LHC are the lightest squarks or the gluino, which will invariably decay in cascades. If R-parity is conserved, the final products must include the lightest supersymmetric particle, which is the lightest neutralino in mSuGra. Ironically, the lightest neutralino will not be directly detected since its interactions are so weak that it will simply fly through the LHC detectors. However, balancing energy and momentum event by event experimentalists will be able to determine its mass and even its couplings with reasonable precision. If this happens, we will be able to assess the portion of dark matter that is made of neutralinos. Even in the lack of detecting supersymmetric particles the LHC will be able to constrain supersymmetry by searching for Higgs bosons. Within the MSSM, for example, there is a strict upper limit of the mass of the lightest Higgs boson and the LHC will be able to probe this limit. If the MSSM is realized in Nature the LHC has to find a Higgs boson below this limit.

Contributions from exotic particles are expected to be important in the case of rare decays. These are decay modes which are forbidden at the lowest order in perturbation theory due to the violation of various quantum numbers, but may proceed through higher order channels mediated by virtual loops. The probability of such decays is hence quite low, however this

means that each higher order diagram will have a greater relative impact. One example is the rate of the rare decay $b \rightarrow s\gamma$, which is considerably less than typical hadronic decay rates. Since supersymmetric contributions to $b \rightarrow s\gamma$ depend on super-partner masses and couplings, the experimental measurement can put bounds on acceptable super particle mass ranges.

There are precision low energy measurements which are sensitive to physics beyond the standard model. Via their quantum fluctuations supersymmetric particles have an influence on various observables we can measure. However this effect is typically suppressed by the particle's mass and relevant couplings and hence is often minuscule. Nevertheless, even minuscule effects can be detected by extremely precise measurements. One such measurement of importance is the anomalous magnetic moment of the muon, which is the contribution to the classical magnetic moment due to quantum fluctuations. There is at present a statistically significant, 3.3 σ, discrepancy between the experimental value and that predicted by the standard model. This discrepancy may be one of the first indications of a supersymmetric effect. Assuming that the discrepancy is due to quantum fluctuations of supersymmetric particles implies an important constraint on supersymmetric models. If the mass of the super-partners is very large then the supersymmetric contribution to the anomalous magnetic moment of the muon is negligible. The supersymmetric contribution only makes up the required difference if some of the super-partners are light enough. This is a crucial constraint affecting the detectability of supersymmetry.

4.1 Bayesian inference

Supersymmetric models may be very compelling mathematical constructions, but studying them quickly raises the question: Is supersymmetry a symmetry of Nature? In the spirit of physics the answer to this question is obtained by contrasting supersymmetric theories with experiments. This is done by comparing various predictions of a selected supersymmetric model to experimental data of choice. However, predictions for the outcome of various experiments in the framework of a given supersymmetric model are not unique. They depend on theoretical assumptions about supersymmetry breaking and, within a given scenario, on the free parameters of the Lagrangian defining the model. This creates a serious obstacle to us learning if and how supersymmetry is realised in the physical Universe. While under the circumstances a straight yes or no answer can not be obtained to our original question, we can still answer various questions that provide vital information on supersymmetry. One of these questions is: Can a given supersymmetric model accommodate our experimental data at all? Another question that we can answer is: What are the values of the supersymmetric parameters most preferred by experiment?

We can answer these questions in a completely rigorous manner using statistics (Baer & Balázs, 2003). The theoretical hypothesis that a certain supersymmetric model exists can be quantitatively tested by statistical methods such as Bayesian inference. Within this framework the answers to the above questions come in forms of likelihood maps over the parameter space of the supersymmetric model.

Such a likelihood map defined over the theoretical parameter space quantifying the supersymmetric hypothesis is $\mathcal{L}(D|P)$. It gives the probability density of obtaining a set of experimental outcome D conditional on setting the theoretical parameters to values P. For various statistically uncorrelated measurements, this likelihood can be calculated as the product of likelihoods corresponding to the individual observations. In the case of normally

distributed uncertainties, we have:

$$\mathcal{L}(D|P) = \prod_{i=1}^{M} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\left(\frac{d_i - t_i(P)}{\sqrt{2}\sigma_i} \right)^2 \right). \tag{10}$$

Here d_i represent experimental central values of measurements and $t_i(P)$ are theoretical predictions that depend on the parameter values for the given experiments. For each observations the statistical and systematic experimental and theoretical uncertainties are quadratically combined into a single uncertainty σ_i. The number of observations is assumed to be M. For measurements that provide exclusion limits, such as searches with negative results, the likelihood is typically given by an error function instead of the above Gaussian form.

The likelihood map constructed over the whole parameter space quantitatively answers the question of whether the supersymmetric model accommodates the experimental data. A low likelihood function over the whole parameter space clearly indicates that the theory is in conflict with the considered experiments. If the value of the likelihood function is sufficiently high somewhere in the parameter space that means the parameter region can accommodate the given supersymmetric model.

While the likelihood function tells us the probability of the experimental outcome for a given set of parameters (under the supersymmetric hypothesis) it does not necessarily tell us the values of the supersymmetric parameters preferred by the data. This is because the latter is given by the conditional probability $\mathcal{P}(P|D)$ which, according to probability theory, is related to $\mathcal{L}(D|P)$ as

$$\mathcal{P}(P|D)\mathcal{P}(D) = \mathcal{L}(D|P)\mathcal{P}(P). \tag{11}$$

The above relation is Bayes' theorem and it tells us that the joint probability of both D and P happening simultaneously can be expressed in two equivalent ways involving the probabilities $\mathcal{P}(D)$ and $\mathcal{P}(P)$ describing that D and P happens separately. Armed with this understanding we can calculate $\mathcal{P}(P|D)$ as the function of $\mathcal{L}(D|P)$ and give an answer to our second question regarding the preferred values of the parameters.

When aiming to infer parameter values from data by calculating $\mathcal{P}(P|D)$, we typically view the probability density $\mathcal{P}(P)$ as prior information on the parameters and refer to $\mathcal{P}(P)$ as the prior. In the same spirit the probability density $\mathcal{P}(P|D)$ itself is usually called the posterior, describing the inferred values of the parameters after information on the data was folded in. But how does $\mathcal{P}(P|D)$ "describe" the preferred value of a given parameter? The answer is clear when we get the rest of the parameters out of the way by integrating over them

$$\mathcal{P}(p_i|D) = \int \mathcal{P}(P|D) \prod_{i \neq n=1}^{N} dp_n. \tag{12}$$

In technical terms this is called the marginalization of the posterior to a given parameter p_i, and N is the number of parameters quantifying the hypothesis. The function $\mathcal{P}(p_i|D)$, when interpreted as the function of a single variable p_i, gives the probability distribution of the parameter. The posterior can also be marginalized to two parameters

$$\mathcal{P}(p_i, p_j|D) = \int \mathcal{P}(P|D) \prod_{i,j \neq n=1}^{N} dp_n. \tag{13}$$

Since $\mathcal{P}(p_i|D)$ is the probability density of the parameter p_i, integrating it over the full allowed region of the parameter gives unity

$$\int \mathcal{P}(P|D) \prod_{n=1}^{N} dp_n = 1, \qquad (14)$$

expressing the fact that the value of the parameter should reside somewhere in its allowed region. This condition can be used to eliminate $\mathcal{P}(D)$ from the calculation of marginalized posterior distributions.

Once the posterior distribution is known we can determine credible intervals for each of the parameters. We define a credibility region $\mathcal{R}_x$ for a parameter p_i by the minimal sized parameter region supporting x percent of the total probability:

$$x = \int_{\mathcal{R}_x} \mathcal{P}(p_i|D)\, dp_i. \qquad (15)$$

In plain terms, a 68 percent credibility interval is the minimal parameter region that contains 68 percent of the area under the posterior distribution. This region gives the location of the parameter at 1-σ certainty. Combined credibility regions over multi-dimensions of the parameter space can be similarly defined as the minimal region satisfying

$$x = \int_{\mathcal{R}_x} \mathcal{P}(p_i, p_j|D)\, dp_i\, dp_j. \qquad (16)$$

The likelihood function can also be projected to one of its variables by forming profile distributions. These are constructed by finding the maximum likelihood hypersurface

$$\mathcal{L}_{max}(D|p_i) = \max_{p_1,\dots,p_{i-1},p_{i+1},\dots,p_n} (\mathcal{L}(D|P)). \qquad (17)$$

and projecting this to one (or more) of the parameters (or to a function of the parameters). These profile likelihoods highlight the model regions where the likelihood is highest.

The numerical construction of a comprehensive likelihood map over a supersymmetric parameter space is required for the evaluation of the marginalized posteriors. This can be a challenging task and its essence is the satisfactory sampling of the likelihood function. The most efficient sampling methods focus on the parameter regions where the likelihood function is the highest. One of the simplest of these methods is the construction of a Markov chain. A Markov chain samples the parameter space in proportion with the value of the likelihood function. High likelihood regions are sampled proportionally higher using the Metropolis-Hastings algorithm. The chain is truncated when some convergence criterion is met. The power spectrum of the chain can be used to show when it is approaching the behaviour of a purely random walk, after which the sample is not gleaning much new information. Parameter regions with several local maxima typically require multiple chains to achieve the most faithful sampling.

In the Bayesian approach, in order to make statistical inferences about the parameter values we must include their prior probability $\mathcal{P}(P)$. Different choices of prior can lead to vastly different inferences. If no prior information is available about a given parameter the choice of a uniform prior appears to be justified. This choice is not trivial however. Should the prior, representing no a priori information, be uniform in the parameter itself or in some function

of the parameters? The answer to this question is not obvious. Present practice in the field of supersymmetric phenomenology prefers two types of priors: uniform in the parameter itself or in the logarithm of the parameter. The latter reflects the lack of prior information on the order of magnitude of the parameter.

5. Likelihood analysis of the next-to-minimal supergravity model

In this section we present our findings for the viability of the NmSuGra model. Due to the complexity of the calculations and the number of observables involved, our numerical program comprised of a number of different codes compiled together, communicating via a sampler implementing the Metropolis algorithm. This code is tasked with selecting points in the parameter space and calculating the likelihood of the results collected. Each chosen set of input parameters is passed to the spectrum generator, NMSPEC (Ellwanger & Hugonie, 2007), which produces the input required for the other codes to calculate their respective observables. These calculations, including the relic density, the anomalous magnetic moment of the muon (Δa_μ), and various quantities related to b-physics, are performed by MicrOMEGAS and other codes in NMSSMTools (Belanger et al., 2006).

For the free parameters of the NmSuGra model, we scan over the following ranges: $M_0 < 5$ TeV, $M_{1/2} < 2$ TeV, -3 TeV $< A_0 < 5$ TeV, $\tan \beta < 60$, $10^{-5} < \lambda < 0.6$. The positive choice of λ ensures that the sign of the superpotential μ parameter is positive as this is strongly favoured by certain measurements including Δa_μ and the branching ratio $b \to s\gamma$. Seperate scans were performed using linear priors and logarithmic priors in order to gauge the dependence of the results obtained on the prior chosen. As the features of these plots are similar, we display only the linear priors, and simply explain the consequences of choosing a logarithmic prior.

We begin by plotting the profile likelihood for the input parameters of NmSuGra in Figure 1. Our first observation is simply that the magnitudes of the dimensionful parameters (M_0, $M_{1/2}$, and A_0) appear to be constrained strongly by experiment. This is due largely to the anomalous magnetic moment of the muon, δa_μ (Balázs & Carter, 2008). In general, as M_0 or $M_{1/2}$ increase, the masses of super-particles increase, and supersymmetric contributions to δa_μ decrease. Thus higher super particle masses incur a penalty to their likelihood up to a limit of 3.3 σ corresponding to $\Delta a_\mu = 0$. In general, the combination of Δa_μ and Ωh^2 defines many of the general features of the likelihood distribution. The likelihood is highest at relatively low values of M_0 and $M_{1/2}$ where light super particle masses help satisfy both constraints.

Figure 2 shows the marginalised posterior probability of the input parameters. The contrast with Figure 1 is considerable, particularly for M_0 and $\tan \beta$. This is caused by the so-called volume effect. The marginalised posterior represents the cumulative probability corresponding to each input value because of the integration in equation (12). As a result, a low average likelihood can be compensated for by simple spatial expanse. Thus, despite the approximately 3σ penalty imposed on the likelihood by δa_μ, high M_0 values may still be reasonable. The large volume elevating the posterior for high values of M_0, and contributing to the sharp peak at $\tan \beta \approx 50$ is called the focus point, shown clearly in Figure 3. Conversely, low values of $\tan \beta$ severely constrict what regions of parameter space are permissible, so the posterior is low in spite of the high likelihood of the viable regions. Figure 2 does not imply any upper bounds on M_0, unlike $M_{1/2}$ and A_0, and unlike what the profile likelihoods suggest. Unfortunately, as M_0 directly influences the masses of sparticles, this may decrease the chances of colliders being able to produce them.
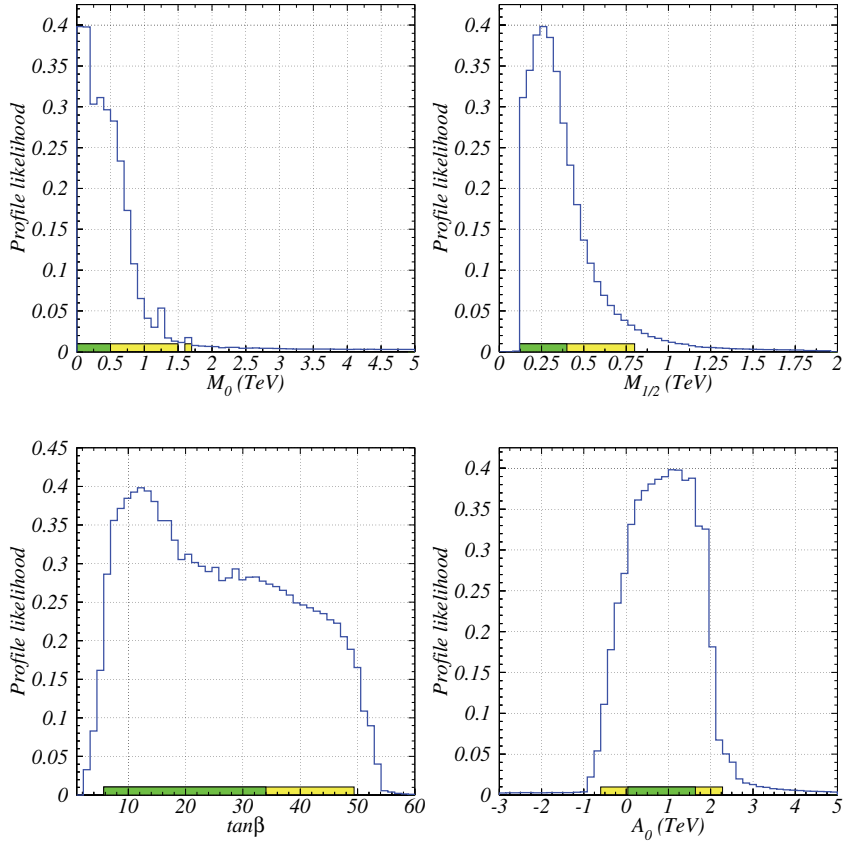
Fig. 1. Profile likelihoods of NmSuGra input parameters.

Naturally, we must keep in mind that these results are contingent on our choice of prior, and in changing this we may have to reconsider what inferences we draw. Logarithmic priors are agnostic to scale, making all orders of magnitude within the scan limits equally likely. Consequently, assuming logarithmic priors instead of linear can mitigate the probability enhancement that the larger volume at high parameter values entails. Certainly, these priors alleviate the potentially problematic high posterior at high M_0.

Our results in Figure 3 are useful for determining the properties of dark matter that we may anticipate for this model. Considering the top-left panel of Figure 3, different annihilation mechanisms can be identified from the structure of the posterior in parameter space. Looking towards the left, when the common scalar mass M_0 is sufficiently low, some scalar particle (generally the stau) becomes lighter than the lightest neutralino, which is forbidden under our assumption of electrically neutral dark matter. However, as the mass difference between these particles decreases the rate of mutual co-annihilation swiftly increases. Hence there is a narrow band of viable parameter space close to the left hand edge. This edge shifts however with the variation of the remaining parameters, so the marginalised posterior in M_0 and $M_{1/2}$ becomes smeared. Higgs resonances, where the lightest neutralino resonance-annihilates to a Higgs boson, occur along thin hypersurfaces where any of the Higgs masses are twice the
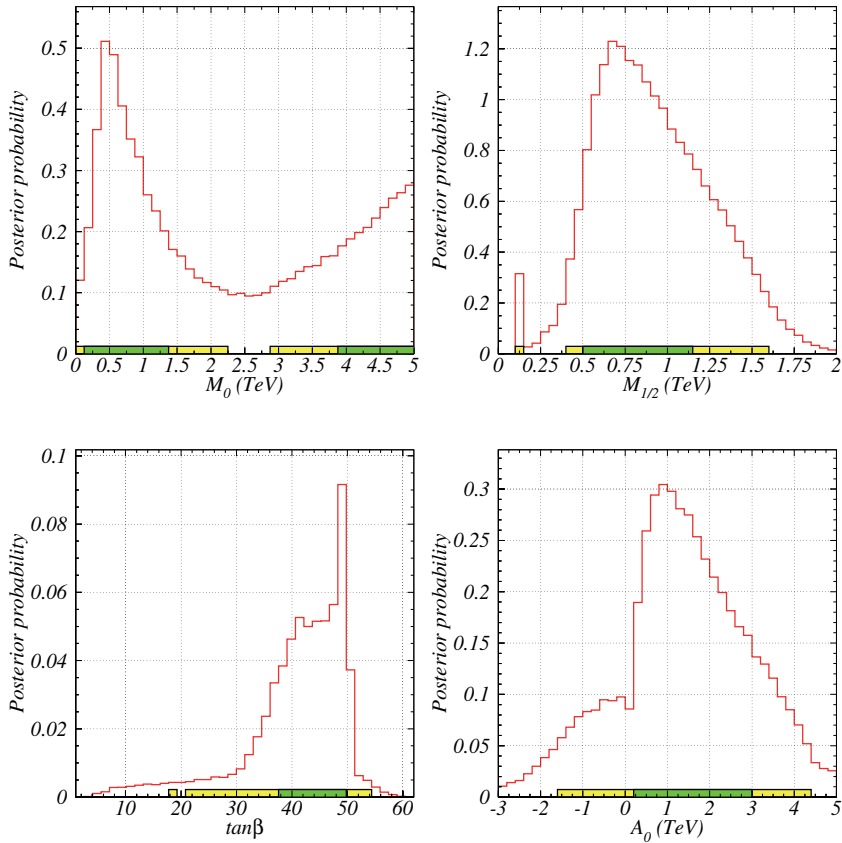
Fig. 2. Posterior probabilities with linear priors marginalised to individual input parameters.

neutralino mass. They can create high likelihood values in otherwise uninteresting regions. Their impact on the posterior can be considerable due to the volume of these Higgs funnels. Along right-hand side of the plot, towards higher M_0 values, we find the focus point region. Here the lightest neutralino is largely higgsino, and it co-annihilates with charginos and other neutralinos. The choice of logarithmic priors typically suppresses the posterior both in the Higgs resonance and focus point regions.

On the top left frame of Figure 3 the diagonal black curve shows the estimated reach of the LHC for 100 fb^{-1} luminosity at 14 TeV center of mass proton-proton collisions. It is evident that with 100 fb^{-1} the LHC will be able to cover the stau-neutralino co-annihilation and part of the Higgs resonance corridors. But the LHC will stop short of fully exploring the high mass Higgs resonance annihilation region and the focus point at high M_0. To put this reach into context: LHC experimentalists celebrated the collection of data corresponding to the first fb^{-1} on June 16 of 2011. This luminosity was collected at half of the design energy, 7 TeV center of mass (3.5 TeV proton beams). Presently 165 pb^{-1} of this data is analysed in the context of the most constrained supersymmetric models (ATLAS collaboration, 2011). This data enables the LHC to reach up to about $M_{1/2} = 450$ GeV at $M_0 = 100$ GeV and to about $M_{1/2} = 240$ GeV
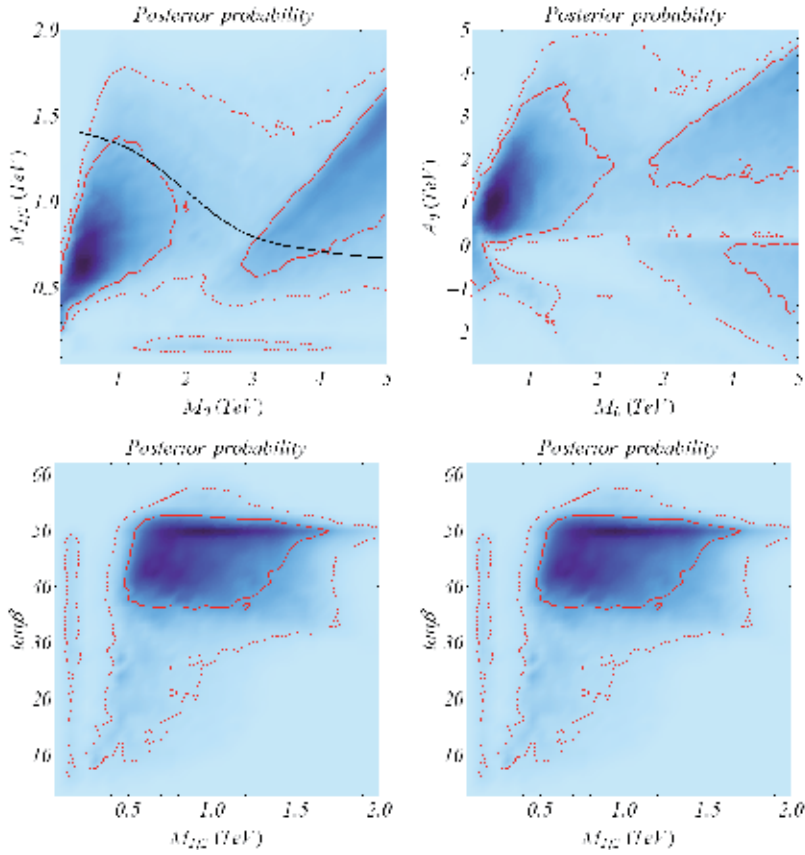
Fig. 3. Posterior probabilities with linear priors marginalised to pairs of input parameters. Various mechanisms for the annihilation of dark matter can be identified from the structure of the posterior distributions. On the top left frame the diagonal black curve shows the estimated reach of the LHC for 100 fb^{-1} luminosity.

at $M_0 = 1000$ GeV. As the top left frame of Figure 3 shows the LHC just started to probe the 68 % confidence level region of the NmSuGra model.

Figures 4 and 5 show, respectively, the profile likelihoods and posterior probabilities of certain interesting super-partner masses. Again, the volume effect skews the posterior probability away from the profile likelihood toward higher values of M_0 and $M_{1/2}$. Although the likelihood can be high at low mass values, these occur only in a very confined region of parameter space. The volume effect is particularly noticeable in the stau mass which closely mimics the behaviour of M_0. Indeed, as the lightest stau is responsible for neutralino annihilation at low M_0 values, and largely responsible for the value of δa_μ, the stau mass accounts to a large extent for the likelihood distribution of M_0. The gluino and stop masses are those of most relevance to detection at colliders, so it is promising that their masses are unlikely to be arbitrarily high (in this model at least). The mass of the lighest neutralino is less on average than the others in part simply because it must be the lightest supersymmetric particle. High neutralino masses generally incur penalties from δa_μ.
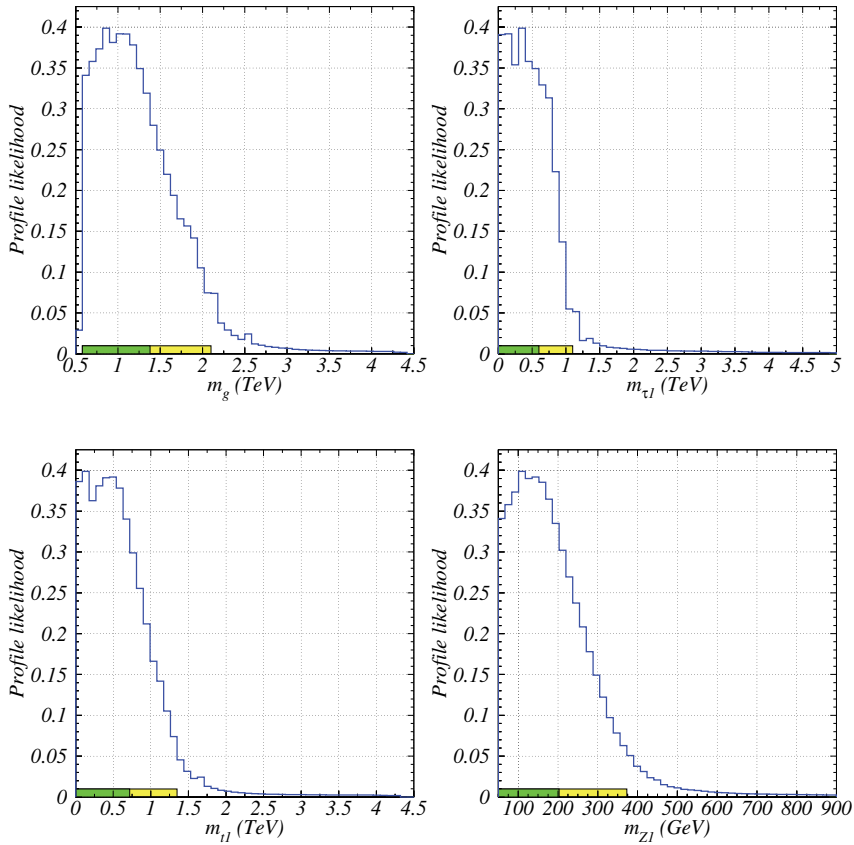
Fig. 4. Profile likelihood distributions for the masses of certain supersymmetric particles, the gluino (top-left), lightest stau (top-right), lightest stop (bottom-left), and lightest neutralino (bottom-right).

5.1 Detection of supersymmetry

The supersymmetric particle that is by far the most likely to be produced at a hadron collider will be one of the lightest strongly interacting super-particles, either the lightest squark (a stop in mSuGra) or the gluino. This particle will then decay in a chain that (assuming R-parity) must contain the lightest neutralino, and will hopefully lead to a discernable signal. Thus, to estimate the capability of a hadron collider to detect supersymmetry, we use the mass of the lightest stop and the gluino. Due to the similarity between the two models, we may apply results already obtained for the MSSM (Baer et al., 2003), assuming an eventual luminosity of $100 \ \mathrm{fb}^{-1}$.

The masses of the stop and the gluino weakly depend on A_0 and $\tan \beta$, so their contours may be plotted to a good approximation as lines in the $M_0 - M_{1/2}$ plane. As shown in the first frame of Figure 3, this contour appears to enclose a large proportion of viable parameter space, which is good news for detection prospects.

It should be clear, of course, that particle mass is a fairly crude indication of the detectability of any given point in parameter space. The production of the gluino and squarks in a
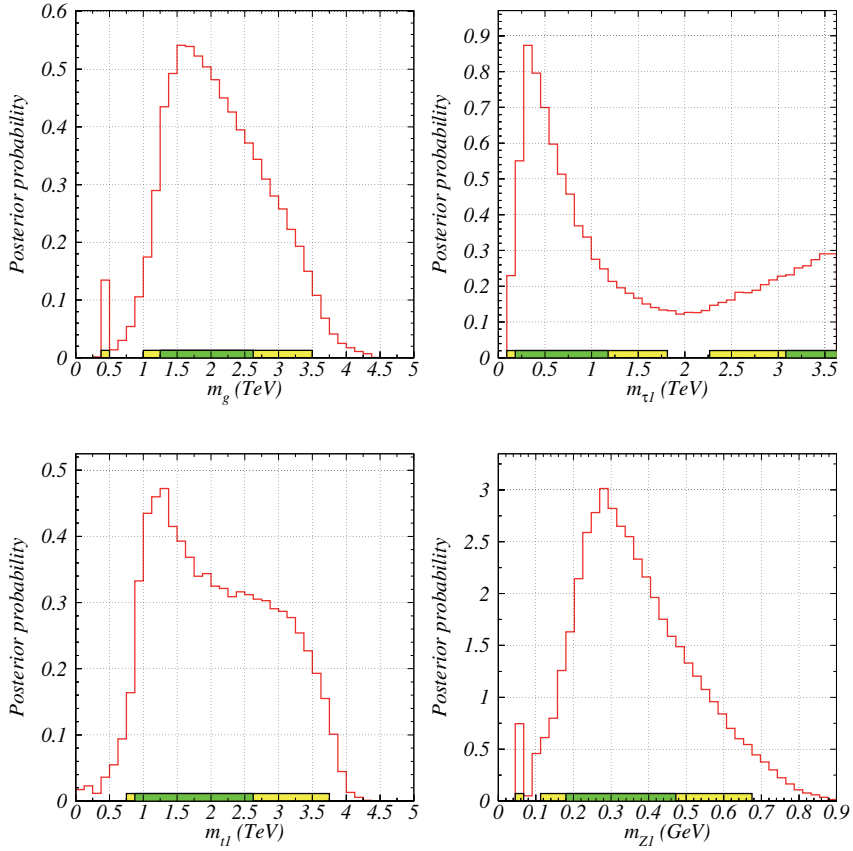
Fig. 5. Posterior probabilities with linear priors marginalised to selected super-particle masses.

hadron collider will be determined by the kinematics, so mass is obviously crucial. However, the significance of the signal (whether an exotic signal can be in fact be verified), depends sensitively on the details of the spectrum, and may be much more difficult to achieve than mere production. However, a full simulation of collider events and detection is a highly complex task, and unfeasible to do for every point in parameter space, so some simplification must be used.

In Figure 6 we show the potential for the direct detection of neutralinos from existing and upcoming experiments. Current exclusion limits are not extremely potent. However, large amounts of viable parameter space seem to be on the cusp of detectability. These points reside mainly in the focus point region. Rather fortuitously future direct detection experiments are able to cover much of the parameter space that will not be accessible to colliders. Based on our assumptions, the bulk of viable parameter space will be covered by either imminent collider or direct detection experiments. We may attempt to estimate the actual chances by quantifying the probability contained within the contours compared with the total amount residing in all parameter space. For the collider reach, we must integrate the posterior probability over all regions in parameter space satisfying either the stop or the gluino mass limit. Doing this

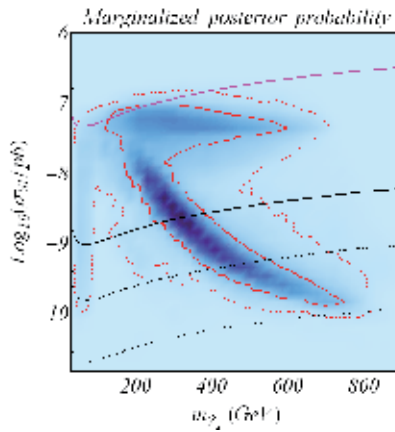Fig. 6. Marginalised posterior probability in terms of the mass of the neutralino and the spin-independent contribution to the neutralino-neucleon cross-section. The present experimental limit is shown as a solid magenta line. The other limits, in order of increasing strength, represent the projected 25 kg, 100 kg, and 1000 kg CDMS reach (Ahmed et al., 2009; Akerib et al., 2006).

calculation with linear priors, we obtain a chance of 57 % that one of these super-particles will be produced at the large hadron collider (assuming, naturally, that our chosen theory is correct). Also encouragingly, with logarithmic priors our chances increase substantially due to the severe disfavour the focus point receives. Including points accessible to 1 ton CDMS, this estimate becomes 99.7 %, all but guaranteeing the eventual discovery of this model if it exists (Balázs & Carter, 2010).

We close our discussion of supersymmetry detection with remarks on gravitational waves. This is because cosmological first order phase transitions may leave their imprint on gravitational waves in the early Universe, and the existence of supersymmetry may imply various such first order phase transitions. The breaking of supersymmetry, for example, may involve more than one metastable vacua. Relaxing via first order phase transitions these vacua may generate a background of gravitational waves. These gravitational waves, in turn, provide an insight into supersymmetry breaking. Supergravity motivated models, such as NmSuGra, are based on a unification of all forces near 10^{16} GeV. Similarly to supersymmetry breaking the breaking of the unified gauge group may be accompanied by gravitational waves. If detected, these gravitational waves can support unification. While supersymmetry provides a dynamical explanation for the spontaneous breaking of electroweak symmetry it may similarly explain the matter-antimatter asymmetry of the Universe via electroweak baryogenesis Balazs et al. (2005). The latter requires the electroweak phase transition to be strongly first order. A strongly first order phase transition generates gravitational waves, the observation of which may shed further light on the mechanism of electroweak baryogenesis and the role of supersymmetry therein.

6. Conclusion

According to our analysis, the prospects for the detection of some supersymmetric particle are quite favourable, at least insofar as supergravity models are concerned. Our results show that the bulk of viable parameter space lies within the potential eventual reach of the LHC

or direct detection experiments even for linear priors, and for logarithmic priors the odds are significantly better. The chances for finding the Higgs boson (again assuming something like supergravity is correct) are similarly promising.

As anticipated, the phenomenology we have observed for the constrained next-to-minimal supersymmetric standard model is not markedly different to what has been obtained for the constrained MSSM. In NmSuGra there are extended neutralino and Higgs sectors. This implies a potential singlet component for the lightest neutralino and Higgs, which could complicate detection attempts. We determine, however, that the presence of these novel features is very limited in the parameter space we explore.

As experimental limits and precision continue to improve, viable regions in parameter space for candidate models will become further constrained, and our ability to discriminate between them should increase. The detection of new particles would be particularly helpful, and between the Large Hadron Collider and dark matter detection experiments the chances for this seem quite reasonable.

7. References

Abazajian, K. N. et al. (2009). The Seventh Data Release of the Sloan Digital Sky Survey, *Astrophys. J. Suppl.* 182: 543–558.

Ahmed, Z. et al. (2009). Search for Weakly Interacting Massive Particles with the First Five-Tower Data from the Cryogenic Dark Matter Search at the Soudan Underground Laboratory, *Phys. Rev. Lett.* 102: 011301.

Akerib, D. S. et al. (2006). The SuperCDMS proposal for dark matter detection, *Nucl. Instrum. Meth.* A559: 411–413.

Amaldi, U., de Boer, W. & Furstenau, H. (1991). Comparison of grand unified theories with electroweak and strong coupling constants measured at LEP, *Phys. Lett.* B260: 447–455.

Amsler, C. et al. (2008). Review of particle physics, *Phys. Lett.* B667: 1–1340.

ATLAS collaboration (2011). Search for squarks and gluinos using final states with jets and missing transverse momentum with the atlas detector in sqrts=7tev proton-proton collisions, *http://cdsweb.cern.ch/record/1356194/* .

Baer, H. & Balázs, C. (2003). Chi**2 analysis of the minimal supergravity model including WMAP, g(mu)-2 and b to s gamma constraints, *JCAP* 0305: 006.

Baer, H., Balázs, C., Belyaev, A., Krupovnickas, T. & Tata, X. (2003). Updated reach of the CERN LHC and constraints from relic density, b –> s gamma and a(mu) in the mSUGRA model, *JHEP* 06: 054.

Baer, H. & Tata, X. (2006). Weak scale supersymmetry: From superfields to scattering events.

Balazs, C., Carena, M. S., Menon, A., Morrissey, D. & Wagner, C. (2005). The Supersymmetric origin of matter, *Phys.Rev.* D71: 075002.

Balázs, C. & Carter, D. (2008). Discovery potential of the next-to-minimal supergravity motivated model, *Phys.Rev.* D78: 055001.

Balázs, C. & Carter, D. (2010). Likelihood analysis of the next-to-minimal supergravity motivated model, *JHEP* 03: 016.

Belanger, G., Boudjema, F., Pukhov, A. & Semenov, A. (2006). MicrOMEGAs: Version 1.3, *Comput. Phys. Commun.* 174: 577–604.

Chamseddine, A. H., Arnowitt, R. L. & Nath, P. (1982). Locally Supersymmetric Grand Unification, *Phys.Rev.Lett.* 49: 970.

Clowe, D. et al. (2006). A direct empirical proof of the existence of dark matter, *Astrophys. J.* 648: L109–L113.

Coleman, S. R. & Mandula, J. (1967). All Possible Symmetries of the S Matrix, *Phys.Rev.* 159: 1251–1256.

Dimopoulos, S. & Georgi, H. (1981). Softly Broken Supersymmetry and SU(5), *Nucl.Phys.* B193: 150.

Ellwanger, U. & Hugonie, C. (2007). NMSPEC: A Fortran code for the sparticle and Higgs masses in the NMSSM with GUT scale boundary conditions, *Comput. Phys. Commun.* 177: 399–407.

Fayet, P. (1975). Supergauge Invariant Extension of the Higgs Mechanism and a Model for the electron and Its Neutrino, *Nucl. Phys.* B90: 104–124.

Goldberg, H. (1983). Constraint on the Photino Mass from Cosmology, *Phys.Rev.Lett.* 50: 1419.

Haag, R., Lopuszanski, J. T. & Sohnius, M. (1975). All Possible Generators of Supersymmetries of the S Matrix, *Nucl.Phys.* B88: 257.

Kneib, J. P., Ellis, R. S., Smail, I., Couch, W. J. & Sharples, R. M. (1996). Hubble Space Telescope Observations of the Lensing Cluster Abell 2218, *Astrophys. J.* 471: 643.

Kolb, E. W. & Turner, M. S. (1990). The Early universe, *Front. Phys.* 69: 1–547.

Komatsu, E. et al. (2011). Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation, *Astrophys. J. Suppl.* 192: 18.

Markevitch, M. et al. (2004). Direct constraints on the dark matter self-interaction cross-section from the merging galaxy cluster 1E0657-56, *Astrophys. J.* 606: 819–824.

Martin, S. P. & Vaughn, M. T. (1994). Two loop renormalization group equations for soft supersymmetry breaking couplings, *Phys.Rev.* D50: 2282.

Pagels, H. & Primack, J. R. (1982). Supersymmetry, Cosmology and New TeV Physics, *Phys. Rev. Lett.* 48: 223.

Rubin, V. C. & Ford, W.Kent, J. (1970). Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions, *Astrophys.J.* 159: 379–403.

Wess, J. & Zumino, B. (1974). Supergauge Transformations in Four-Dimensions, *Nucl. Phys.* B70: 39–50.

Matter-Antimatter Asymmetry and States in the Universe

F. L. Braghin

Instituto de Física, Universidade Federal de Goiás, GO

Brazil

1. Introduction

Dirac's interpretation of the negative and positive energy eigenvalues of relativistic free fermions equation became a sound paradigm widely and continuously articulated in particle, nuclear physics and modern astrophysics and cosmology. The foundations of the theories for elementary particles and fields rely deeply in symmetries, including those relating particles and antiparticles. In spite of that, observed cosmic rays (CR), diffuse gamma ray spectra and fluctuations of Cosmic Microwave Backgound (CMBR) do not allow to conclude there is a meaningful amount of antimatter in the Universe. If the exact symmetries of particle physics are considered to give rise to macroscopic scenarios that evolved towards the current Universe a clear contradiction appears with the observed matter-antimatter asymmetry.

Although electric charge $U(1)$ and baryon number $U(1)$ have strict conservation laws, GUT and SUSY-GUT predict baryon nonconservation via proton decay: $p \rightarrow \pi^0 + e^+$ and $p \rightarrow \bar{\nu}_\mu + K^+$, and the current lower bound for these processes is of the order of 10^{34} years (69). From the point of view of quantum field theories, a local hermitian quantum field theory in the Minkowski space with Lorentz covariance produces the same mass, lifetime and magnetic moment for particles and their antiparticles with opposite charges according to the CPT theorem (55; 94). This theorem establishes a sound framework on which (most of) particle and cosmological theories have been worked out. Nevertheless, in the scope of the CPT theorem, mechanisms for baryogenesis can be envisaged since C and CP are broken in Nature, at least by weak interactions. Whereas breaking of CP is invoked very often for describing baryogenesis, CPT might not have been valid anymore in the extreme conditions of the Early Universe, and the consequences can be very rich for the cosmo evolution. Tests of the validity of the CPT theorem become an important task (32), and actually there is evidence of CPT violation from neutrino oscillation (3).

Standard Big-Bang theory (88) predicts an equal amount of baryons and photons produced in the Early Universe which is well reduced nowadays. Estimations based on the standard Big Bang nucleosynthesis (98) taking into acount CMBR fluctuations (90) yield the current value for the ratio of the baryon density to the photon density to be around (93):

$$\omega = \frac{\rho_B - \rho_{\bar{B}}}{n_\gamma} = (6.1 + .3 - .2) \times 10^{-10} \qquad (1)$$

where $\rho_B, \rho_{\bar{B}}$ are the baryon and antibaryon densities and n_γ the photon density. Other similar estimations provide: $\omega = (4.7 + 0.1 - 0.8).10^{-10}$ (97) and $(5.0 \pm 0.5).10^{-10}$ (27). Baryons seem

to represent only nearly 5% of total energy density of the Universe while dark matter would correspond to at least 23% of this total energy density (34). Therefore one of the first natural candidates to constitute dark matter was (hidden) antimatter. However this hypothesis did not show to be really reasonable due to the absence of the corresponding particle-antiparticle annihilations in the regions where dark matter is expected to be. The current bound on matter-antimatter asymmetry is nearly $n_{\bar{B}}/n_B < 3.10^{-6}$ (93) basically from the absence of gamma rays from annihilation processes, and that goes along with observation of cosmic rays. In the last years several high precision cosmic ray observatories outside the atmosphere were built and launched (PAMELA, BESS, AMS-2,PEBS), (diffuse) gamma ray spectra in different energies are also continuously improved (EGRET, FERMI and others).

There are different possible solutions for the contradiction between the fundamental elementary particle symmetries of particles and antiparticles and the apparent absence of antimatter in the Universe: (1) there was at least one reasonably efficient mechanism of producing more matter than antimatter (baryogenesis) , (2) antimatter continuously existed and currently it is somehow hidden (either in some different place(s), domains or state(s)), or (3) antimatter would not have been as stable as matter and it could have been annihilated or decayed in some other way that was not the usual particle-antiparticle annihilation. In particular, mainly (but not only) for the solutions (2) or (3) antimatter may be distributed non homogeneously in the Universe within the so-called "islands" (or domains) of antimatter. The picture of matter dominated Universe is corroborated by estimations on the minimum distance to the nearest antimatter dominated galaxy or cluster that could agree with CMBR fluctuations, and which is expected to be farther than around $10 - 20Mpc$ (31; 57; 92; 93). Within this scenario one usually is left with the problem of efficient baryogenesis in the Early Universe, although the domains of symmetric Universe still might have this size. The most well known mechanism was proposed by Sakharov (79) and it relies in non-equilibrium dynamics and breaking of CP, although in some models it is not enough to produce the expected asymmetry (34). However there are other mechanisms that have been invoked to explain such matter-antimatter asymmetry, mainly for solutions (1) and (2), as discussed in Sections 3 and 4. Among them, one can consider the existence of very dense antimatter objects, such as primordial (anti)black holes (PBH) in larger number than matter objects (anti-black holes). Often, models accounting for any of the three solutions are not compatible with other aspects of Standard Big Bang theory (such as nucleosynthesis) and CMBR fluctuations. However, antimatter still might be somehow hidden from observations, and an eventual realistic estimation of the possible location(s) or states, for example in the form of anti-stars or Galaxies, depends on several hypothesis, being subject of theoretical and experimental investigation. In any case, eventually one can expect to find effects that could prevent the corresponding annihilation of matter and antimatter in different phases of the Universe evolution, such as discussed in Section 4.

The Chapter is organized as follows. In the next Section, a short review of observations related to antimatter in the Universe is presented, mainly referring to cosmic rays and diffuse gamma rays. In Section 3, we discuss the current constraints for the existence of large domains of antimatter that would be expected in a $B - \bar{B}$ symmetric Universe and the usual baryogenesis conditions proposed by Sakharov (79). Some of the most successful scenarios in which the Sakharov conditions work are quoted, besides some other alternative models for baryogenesis. In Section 4, we discuss few different models that can also allow for symmetric baryon-antibaryon Universe. Some dynamical effects that might make possible these less usual scenarios are discussed. Some aspects of relevance for dense antimatter

bound systems, such as dense antistars, are also presented. In the last Section, there is a brief summary of the ideas presented in the Chapter. The amount of works about these subjects is large and this Chapter is not intended to provide a complete exhaustive review of them. Given the large number of possible mechanisms involving production or dynamics of antimatter some of the most often considered will be shortly described. This work was, in part, based in some good reviews on aspects of the antimatter problem available in the literature (15; 28; 34; 36; 38; 54; 78; 84).

2. Observational constraints: results and plans

There are two main types of observations for detecting antimatter in the Universe: gamma rays spectra and cosmic rays (CR). However, as it will be noted, these observations must be considered not only as resulting from the processes involving primary production of antimatter or its annihilation but also from secondary interactions of both antimatter and matter. Besides these observables, we mention other possible signatures of antimatter, such as isocurvature perturbations for which there still are uncertainties (61). Isocurvature fluctuations do not exclude the possibility of antimatter objects in not far distances from us (39; 61). Neutrinos (and antineutrinos) from supernovae are observed in Earth observatories, therefore one could also expect to observe antineutrinos from anti-supernovae with typical flux but seemingly these events have never been observed (39; 77). Gravitational lensing for halo compact objects is considered by the MACHO project (*Massive Compact Halo Objects*). White dwarfs or even black holes can be searched (7) and the association of these observations with gamma rays can also provide information on the presence of antimatter in this region of the Milk Way.

For the high precision CMBR fluctuations and diffuse gamma ray spectrum several observatories are working and being planned, such as: WMAP (103), Cosmic Background Imager (29), Planck mission (44) and FERMI (43). Planck mission, for example, will measure the CMB with very high angular resolution and high sensitivity, improving the COBE resolution. We just mention further FERMI observatory (Fermi Gamma-ray Space Telescope, formerly GLAST) that is a high precision observatory to analyse high energy events in the Cosmos (43): black hole dragging matter, emission of jets with high speeds, signals of dark matter, gamma ray bursts, origin of cosmic rays, pulsars, among others. Certainly the corresponding gamma ray spectra is expected to provide valuable information on the contribution of antimatter components via annihilation processes in such high energy events. Differences on the spectrum with respect to previous observations have appeared but deserves further investigation (2; 108). It is assumed that proton-antiproton annihilation via π^0 is the most important baryon annihilation process for the cosmic diffuse gamma spectrum (91). Given that there is an average of four γ per $p - \bar{p}$ annihilation and two γ from $e^+ e^-$ annihilation (with $E = 0.511\text{keV}$ in the restframe) the resulting spectra would be quite peculiar and quite strong (both lines: from $e^+ e^-$ and π^0 decay due to $p - \bar{p}$ annihilation) and this is not observed. One obtains an upper limit on the antimatter/matter fraction to be around 10^{-15} in Galactic molecular clouds, 10^{-10} in the Galactic halo, and 10^{-5} within our cluster (75). However, the line of $E = 0.511\text{keV}$ ($e^+ e^-$) is quite bright and visible from the Galactic center, from Galactic bulge and possibly even in the halo (76; 100; 101). It is not settled whether and to what extent they are fingerprints of antimatter in these places. Diffuse gamma rays and CMBR fluctuations help to put a constraint on the minimum size of our matter dominated region out of which it could appear an antimatter dominated "island", and it is nearly 20 Mpc (31). Taking into account the pressure of such radiation, with dragging

effect from diffusion of protons and antiprotons, the scale of matter dominated region in the Universe would be extended to 1 Gpc depending on the model considered (39). Unless new dynamic effect(s) provide a well defined separation of large antimatter domains during the Universe evolution these domains are rather ruled out as considered so far. However, several of the observations cannot be considered as proofs of antimatter absence, Earth fortunately seems to be deep in a matter dominated region that prevents antimatter (or, to some extent, its fingerprints) to reach us.

On the other hand, the gamma ray spectra coming from annihilation of antimatter in astrophysical objects (or anti-objects themselves) are easier to interpret. Annihilation of proton flux towards an anti-object is given for bulk and surface regions in Ref.(15) resulting in very high luminosity events. Punctual annihilation of antimatter (either in the form of free propagating antiparticles, small objects or antistars) in gas clouds as well as collisions of antistars with stars of different types were also considered in (15). By examining the collision of two galaxies, the presence of matter-antimatter can be detected by annihilation which emits the characteristic gamma rays yielding a maximum relative amount of antimatter in our Galaxy of the order of 3.10^{-6} (93). These last events can be extremely energetic and maybe providing gamma bursts.

2.1 Cosmic rays

Cosmic rays CR can provide one of the direct fingerprints of antimatter structures in the Universe, in spite of the interaction with the interstellar medium (ISM). In any case, the need of observations outside the atmosphere is undeniable. Primary sources have been proposed to be galactic and extragalactic. Firstly one can expect that antimatter objects can emit antiparticles (antiprotons, positrons and eventually anti-nuclei) in the same proportion as matter objects emit particles, however interaction with ISM can well contribute for the CR spectrum as well as other primary sources: evaporation of micro BH (64; 106) or annihilation of dark matter from the galactic halos (eg. neutralinos) (18). On the other hand antiprotons and positrons can be produced in secondary reactions in the ISM with known cross sections. Some of the most relevant processes for understanding all the antimatter components in the CR are: (1) diffusion in the magnetic fields of the galaxy and of the Sun, (2) reacceleration, after the movement of scattering centers, (3) antiparticles can loose energy by Coulomb scattering or ionization in the interstellar medium (ISM), (4) convection can expel antiparticles from galactic disc or solar system, (5) they can simply annihilate on ISM, usually with H or He (67). The influence of several of these processes on the resulting spectra depend on theoretical predictions and observational astrophysical data. Therefore although CR provides an important source of information about antimatter in the Universe its understanding must include several other aspects. Protons from usual matter can form a massive positively charged BH which produces strong enough electric field to generate pairs of e^+e^-. It should follow a strong absorption of the electrons and ejection of positrons becoming a source of antimatter (16). Dark matter scalar singlet model, for example, predicts annihilation of scalar dark matter into positrons and antiprotons that might be in part those observed in high energy CR (48). although the mass of such scalar is not measurable not even in LHC. Galactic halos might contain substructures (clumps) and may enhance dark matter annihilation. Basically, the current upper bounds of antimatter from CR can be considered to be nearly 10^{-4} for $\bar{p}/p$ and 3.10^{-8} for $\bar{He}/He$, which might settle nearly the scale of maximum amount of antimatter in the galaxy. There are several on going projects intended to clarify the antimatter components of CR (positron / antiproton/

antinuclei): PAMELA2006 (72; 74), AMS02-2009 (11), BESS (19), PEBS-2010 (46), GAPS-2013 (65).

PAMELA (A Payload for Antimatter Matter Exploration and Light-nuclei Astrophysics) (72; 74) offers high statistics and sensitivity for the composition of the spectrum at high energies. Positron flux are measured with energies from 50 MeV to 270 GeV (previous limits were $0.7 - 30$ GeV), antiproton flux from 80 MeV to 190 GeV (previous limits $0.4 - 50$ GeV) and antinuclei ($\bar{H}e/He$) with flux of nearly 10^{-8} (previous limit about 10^{-6}). Besides that, matter will be observed in a wider range of energies as well.

AMS02 collaboration is a worldwide collaboration (11; 89) that will search for positron, antiproton, anti-helium and heavier antinuclei (up to anti-iron) with energies up to 1 TeV with higher statistics and sensitivity than AMS.

BESS (Balloon-borne Experiment with Superconducting Spectrometer) (19) offers a much lower cost than space missions by long duration polar ballon flights. It is suitable to detecting low energy antiprotons, antideuterons and antihelium (19; 71; 82). However it is GAPS that offers the best apparatus for detecting antideuteron (65).

The relative antiproton flux $\bar{p}/p$ observed in different satellites/observatories is presented in Fig. 1 (figure from Ref. (66),*Morselli, Moskalenko, Status of indirect searches in the PAMELA and FERMI era* Proceedings of Science, PoS, *http://pos.issa.it (IDM2008) 025)*). Antiproton flux is similar from the different observations (BESS, CAPRICE, MASS, PAMELA, IMAX, HEAT) In solid lines lower and upper limits on the contribution for pure secondary component from Ref. (95). Secondary interactions are overestimated by usual codes (75; 87; 95). Therefore the peak in the low energy around 2GeV presents the behavior described by secondary interaction component of the ISM although it is probably not described only by these interactions.
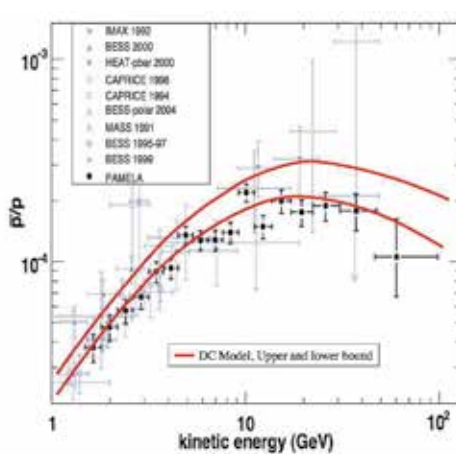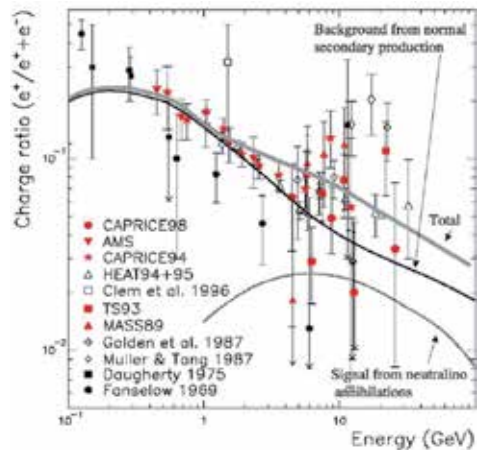
(a) Figure 1   (b) Figure 2

Fig. 1. Antiproton flux from the different observations (BESS, CAPRICE, MASS, PAMELA, IMAX, HEAT) and with estimations for secondary ISM interactions from (95). Figure from (66). **Fig. 2**: The relative positron flux, $e^+/(e^+ + e^-)$, from Ref. (75) with data from several experiments and two theoretical curves (model with pure secondary interaction in ISM (in black solid line) and a model in which neutralino annihilates producing positrons)

The relative positron flux, $e^+/(e^+ + e^-)$, from different observations, is shown in Fig. 2 (figure from Ref. (75), *Picozza, Morselli, (2008)* Journal of Physics: Conference Series *120,042004, IOP Publishing*), where the total flux is in grey solid line. Since these particles are light they must loose energy considerably in interaction with interestellar matter. On the other hand positrons might also originate from secondary interaction in such interstellar medium (from $\pi+$ and K^+). Two theoretical curves are included: model with pure secondary interaction in ISM (in black solid line) and a model in which neutralino annihilates producing positrons (in dotted line (14)) There is a slightly reduced decrease of the total flux for energies around 10GeV.

From the different observations in the last years the upper limit of the relative flux for heavy anti-nuclei to nuclei (Z>2) is around 1 among 10^4 nuclei and the current bound on the ratio of anti-helium/helium is of the order of 3.10^{-7} from BESS-Polar and BESS-TeV (82). These observations already shows an improvement compared to the previous AMS upper bound that was 10^{-6} (10). The absence of light antinuclei (mainly $\bar{He}$) emitted by antistars could be (at least partially) explained by interaction in the matter dominated ISM and eventually dragging by their electric (or magnetic) fields. At a first look, one usually expects nearly the same proportionality of particles and antiparticles observed in CR and the relative abundance of matter and antimatter mainly in the Galaxy, i.e. composing the emitting sources of CR. Following this reasoning Dolgov and Bambi (15) have found an upper bound on the number of antistars in the galaxy considering the total number of stars is of the order of 10^{11}. Besides that, considering the current bound on the ratio of anti-He/He content of the CR one has (15):

$$N_{antistars} \leq 10^3. \qquad (2)$$

It reduces to a extremely small fraction and its relative amount is far from being enough to justify a matter-antimatter symmetric picture of the galaxy.

2.2 Current tests on antimatter

In spite of the evidences for the need of baryogenesis mechanisms it is important to testify further antimatter, the properties and fundamental symmetries we assume to be valid, such as CPT and its consequences. Not only the current validity of these symmetries must be verified but also their behavior along the evolution of the Universe should be investigated, in particular in the Early Universe, when the extreme conditions could be responsible for modifications of matter/antimatter properties. Whereas the first of these aspects can usually (but not always) be tested in laboratories, the second remains rather an open field that continuously receives more input from theories and observations. The equivalence between inertial and gravitational mass for antimatter (anti-nuclei) is tested in the AEGIS experiment at CERN (4; 40) ("Antimatter Experiment: Gravity, Interferometry, Spectroscopy") through positronium/ antihidrogenium measurements. Different tests on CPT will be done in CERN: by ALPHA experiment (6) will consider spectral lines of anti-hydrogen atoms, ASACUSA will test the hyperfine splittling (12), ATRAP will perform accurate laser spectroscopy on anti-hidrogen (13).

An interesting recent test on CPT has appeared in neutrino physics from MINOS experiment (3). Oscillations of neutrino and antineutrinos have been found to be compatible with a violation of CPT although one expects to have more precise results from forthcoming experiments.

Some other tests on antimatter interactions and bound states will be done by FAIR-GSI (42).

Although there is no available way of testing some further issues we mention that besides the possible CPT breaking that could have modified deeply the expected dynamics of the Early

Universe, some other modified gravitation theories which lead to effective repulsion between matter and antimatter in the Early Universe have also been envisaged (50). There are several theories yielding antimatter behavior different from matter, just as example we quote Ref. (86) with theories of supergravity for which Kaluza-Klein graviscalar and gravivector components may provide different couplings between matter/antimatter.

3. Cosmological baryogenesis

CMBR fluctuations and diffuse gamma ray spectrum are compatible with matter dominated Universe. If the Universe presented the same amount of matter and antimatter the most probable realisation would appear with whole domains of antimatter (anti-galaxies, anti-clusters). However, the best established theoretical estimations for this scenario indicates that if there were boundary walls of matter and antimatter dominated regions ("islands") they would have strong annihilation effects in the borders already in the phase of hydrogen recombination that would distort the CMBR anisotropies and polarization (31; 68; 93). After recombination it would be difficult to prevent annihilation for different domains of matter and antimatter. If the Universe contains a sizeable (even if not large) baryon excess one is left with the problem of understanding the mechanisms that yielded the corresponding symmetry breakings of particle Physics. Given that at very high temperatures, and more generally high energy densities, the symmetries usually have the tendency of being rather restored (58), even if this is not always expected (20), a scenario that copes such opposite expectations may become a still more difficult task. In this section we discuss aspects of particles field theories and dynamics relevant for the investigation not only of the matter-antimatter asymmetric scenario in the Early Universe but also for different pictures (less asymmetric) of the Universe and antimatter states.

3.1 Sakharov's conditions
The Sakharov's conditions were proposed in the 60's (79) and remain the most well established framework to describe baryogenesis. They assume CPT symmetry along the whole Universe evolution. The conditions are the following:

- Non conservation of baryon number B,

- C , CP and T violation,

- Out of equilibrium environment.

Consider only the first condition that is the essential one for any mechanism. There is no experimental evidence of B non conservation up to 10^{34} years (69), so B number is a conserved quantity in the Standard model Lagrangian although quantum effects might break such symmetry and GUTs also predict its breaking. If C is conserved then a reaction $(X_1 \rightarrow X_2 + B)$ produces a baryon B and the conjugated reaction $(\bar{X}_1 \rightarrow \bar{X}_2 + \bar{B})$ would also take place producing the same amount of $\bar{B}$ over a long period of time. Therefore one needs to break invariance over C to produce (reasonable amount) of B asymmetry. If CPT is considered to be an exact symmetry, the masses of particles and antiparticles are equal $(m_X = m_{\bar{X}}$, $m_B = m_{\bar{B}})$, then an initial baryon asymmetry (from the first two conditions above) would, in thermal equilibrium, be compensated by other processes which would generate pairs of particle-antiparticle so that B would be conserved over a, not necessarily long, period of time. Therefore it is required that during some time B is nonconserved outside equilibrium and when restoring thermal equilibrium this asymmetry remains. Therefore time reversal must also be broken which means that if CPT theorem is intact by the time

of such particle production, CP must be broken as well. If these processes take place out of thermodynamic equilibrium one can expect that the reverse reactions of those breaking B and CP occur with different widths. By the way, one is left with one of the most fundamental links for relating microscopic symmetric world, with the corresponding symmetries, and the macroscopic thermodynamic Universe. Nowadays there are several known mechanisms of CP breaking, although in the Early Universe CP breaking could had taken place by means of mechanisms probably, at least in part, different from those observed in laboratory nowadays (26). However any fraction of matter excess might have been produced by combined mechanisms of spontaneous and explicit breaking of CP (60). It is also plausible that the amount of matter generated by the fundamental processes was simply larger than the amount of antimatter, due to some other mechanism of baryon number violation.

3.2 Some scenarios for Sakharov conditions and other mechanisms

There are many proposals for alternative scenarios to explain the baryon asymmetry in the Universe, we mention few of them in the following and the reader can find more in reviews available in the literature, for example in Refs. (15; 28; 34; 36; 38; 54; 78; 84). In GUT, baryogenesis can be realized from condensation of scalar SUSY-fields in the Affleck-Dine scenario. Scalar superpartners of baryons or leptons (χ) undergo condensation by spontaneous symmetry breaking and their decay involve B-nonconserving processes that can bring a large baryon asymmetry (although this scenario also can favor matter/antimatter domains) (9). The decay of Higgs or gauge bosons X ($m_X \sim 10^{16}$GeV) in GUT might also generate baryon excess (47). Combination of the above processes could be realised, for example through baryogenesis after leptogenesis (45) at the energy scale of $T \sim 10^{10}$GeV. SUSY models can yield baryogenesis resulting from thermal leptogenesis, although usually producing wrong primordial abundance of elements. Some works try to mend this probelm (33). Assuming a SSB of a global $U_A(1)$ gauge symmetry the decay of a Goldstone field can yield baryogenesis in the broken phase (30). However these models easily produce large deviation from isocurvature density fluctuations. Some other models consider interactions violating R-parity which yields an excess of baryons over antibaryons, besides explaining the dark matter with good agreement with current estimations of relative abundance of elements (59).

Some of the processes that could have provided CP- breaking in the Early Universe are not expected to take place anymore nowadays. We remind some mechanisms of CP breaking and the reader will find further details in reviews available in the literature (15; 36; 38; 99). Some CP breaking mechanisms come from an explicit Lagrangian term and others from spontaneous symmetry breaking (63; 99). However this last mechanism yields rather a symmetric universe with equal amounts of matter and antimatter (105). Several of the known mechanisms were proposed within the electroweak theory, sphalerons undergo a tunneling between the degenerated vacuum of different baryonic number yielding B non conservation. Strong-CP violation (96) might contribute to yield sizeable baryon excess if there is large deviation from thermal equilibrium. However the massive Higgs can favor a second order phase transition with small deviation from thermal equilibrium, besides that this CP violation is seemingly weak (36). It is also possible to have dynamical or stochastic CP breaking due to complex scalar field out of equilibrium for which one would have generated a meaningful amount of antimatter (35; 36). A possible solution for the difficulty in producing a correct amount of CP breaking would be TeV - high energy scale gravity (8). In some other models the neutrino mass generation is related to the baryogenesis (81) that also describes dark matter

in a seesaw-like model (51). One has also proposed that although quarks/antiquarks have been created without an asymmetry, dynamics of the early Universe would have found a way of store antiquarks in lumps of color antiquark condensates within (small or not) compact antimatter objects (24; 70).

4. Further ideas

In spite of the (strong) observational evidences discussed above for an asymmetric $B - \bar{B}$ Universe discussed above there still are other possible scenarios, in particular with less asymmetric Universe. For instance considering a complex scalar field, undergoing spontaneous symmetry breaking, and acquiring a vacuum expectation value it provides a mechanism that allows for separating matter and antimatter domains (62). Models for inhomogeneous baryogenesis within Affleck-Dine scenario were discussed in Ref.(37) which give rise to some ways of accommodating antistars from early Universe dynamics considering the decay of scalar baryon condensate χ into quarks (and antiquarks) (37). The order of the phase transition can have strong effects on the resulting distribution of antimatter and it is not known (37). An analysis based in the Coleman-Weinberg potential for the evolution of the χ condensate is given in Ref.(38). In such model, CP is unbroken but B is broken. These scenarios would have to cope with the high energy gamma ray background (31) and with the known "matter" domain wall (107).

Besides the usual picture of antimatter islands discussed above, other related configurations may arise, for instance, antimatter dominated region in the edges of the observed Universe. This scenario could emerge if matter and antimatter interactions have had strongly repulsive interaction in the Early Universe. This scenario is allowed by constraints on CMBR homogeneity. Differently, if the early Universe dynamics allowed for the formation of localized massive antimatter objects (black-holes, B-balls, droplets, nuggets, lumps) they could carry (at least part of) the primordial antimatter (17; 24; 70). Among the localized objects, primordial anti-black holes could be candidates to contain a sizeable amount of antimatter from the Early Universe, therefore, in spite of the difficulties, it would be very interesting to try to distinguish them from matter BH, for example by annihilation processes in the hair. There is no specific limit on the PBH masses, but they might be as small as $10^{-14} M_{sun}$ (solar mass) and reach many times the solar mass. B-balls might also be objects or anti-objects (15), originated from a bubble in the Affleck-Dine mechanism with scalar baryon field χ which decays into baryons/quarks. A different form of bound states of antimatter could be the (anti)lumps formed when quarks and antiquarks were produced in different bubbles in the QCD phase transition (70). They could contain superconducting state anti-quarks (23; 24; 70). Instead of discussing issues very commonly addressed, such as macroscopic hydrodynamics in expanding geometries, we go through some aspects of interest of quantum fields with focus on approximative solutions with qualitative interesting effects. Although the particles are the dynamical variables that constitute the whole Universe, it is an impossible task to perform any calculation or simulation of the Universe evolution, including the formation of structures, considering a time dependent quantum field theory with all the relevant fields and particles. Instead, a hydrodynamic model must be considered in the corresponding curved space time background (31; 102) and the underlying field theories can be considered for the determination of relevant microscopic effects and initial conditions. Even if the initial conditions have shown to not produce relevant effects as discussed in Ref.(31), the microscopic dynamics may introduce further input in the macroscopic scenario. Furthermore, instead of picking up a specific model and analyzing the efficiency of the processes allowed

by their symmetries, we analyze effects that might be present in different theories, and the understanding of the quantitative reach is not addressed. Commonly these results can be implemented with some plausible modification of the former model. Consider a local field theory with fermions (ψ), gauge/vector fields (A_μ, V_μ), scalar field(s) (ϕ, χ). In a curved space-time background with non minimal coupling of gauge and scalar fields to gravitational field one can write the following action (21; 55):

$$S = \int d^4x \sqrt{-g} \left\{ \frac{i}{2} \left[\bar\psi \gamma_\mu \nabla^\mu \psi - (\nabla_\mu \bar\psi)\gamma^\mu \psi \right] - M\bar\psi\psi - a_1\phi\bar\psi\psi - (m^2 + \xi_1 R(\mathbf{x}))\phi^2(\mathbf{x}) \right.$$

$$\left. + R(\mathbf{x}) + \mathcal{L}_{\phi,A_\mu}[\phi,\chi] + \xi R(\mathbf{x})A_\mu A^\mu \right\}, \tag{3}$$

where $\sqrt{-g}$ is the square root of the determinant of the metric, ξ_i are couplings of gravitational field, $R(\mathbf{x})$ is the Ricci scalar, $\mathcal{L}_{\phi,A_\mu}[\phi,\chi]$ is the Lagrangian density of the scalar fields with interactions, including self-interacting terms for scalar fields $V[\phi,\chi]$, and their dynamical equations are not written. In a curved space-time background the Dirac equation for fermions $\psi(\mathbf{x})$ and the corresponding $\bar\psi(\mathbf{x})$ interacting with scalar and vector fields can be written as:

$$\left(i\gamma_\mu(\nabla^\mu + igA^\mu) - M - a_1\phi\right)\psi(\mathbf{x}) = 0, \quad \left(-i\gamma_\mu((\nabla^*)^\mu - igA^\mu) + M + a_1\phi\right)\bar\psi(\mathbf{x}) = 0, \tag{4}$$

where the covariant derivative ∇_μ includes the spin connection (21; 73). In Minkowski space for free fermions the eigenvalues of these equations are $E_0 = \pm\sqrt{k^2 + M^2}$. For the high energy environment of the Early Universe, we assume that the boson fields of this model can acquire a non zero classical value, typical from spontaneous symmetry breakings (55; 80). From these expressions, effective masses can be defined for the bosons and fermions. For example for the fermions (and anti-fermions): $M^* = M - a_1\bar\phi$ in terms of a classical counterpart of the scalar $\phi(\mathbf{x})$. The eigenvalues of the Dirac equation with a classical counterpart of V_μ and scalars ϕ, as external fields, become:

$$E_i^\pm = g_V V_0 \pm \sqrt{(\mathbf{p} - g_V\mathbf{V})^2 + (M_i^*)^2} \tag{5}$$

where the coupling to scalar field was incorporated in the effective mass. In particular little attention has been given to the formation of antimatter and matter-antimatter bound states, (104). Such states hardly would be responsible for absorbing antimatter excess, although they may be envisaged to produce new forms or structures of matter/antimatter eventually coexisting at least for some time in the Universe, maybe also contributing to the radiation spectra. More realistic cases appear when these external fields are not homogeneous quantities, i.e., when $V_0 = V_0(r)$. One can envisage different geometries for the dynamics of the creation of particles and antiparticles such that inhomogeneous configurations can arise that modify the matter-antimatter annihilation ratio or even preventing large annihilation ratio. For the sake of the argument, suppose that dynamics of (strongly interacting) fermions/anti-fermions are such that in a given direction x_i

$$\partial_{x_i}\psi(\mathbf{x}) \sim ik_iF_i(k)\psi(\mathbf{x}), \quad \partial_{x_i}\bar\psi(\mathbf{x}) \sim -ik_iG_i(k)\bar\psi(\mathbf{x}), \tag{6}$$

where F_i and G_i can be functions of momenta and depend on the boson fields. Although one would expect solutions with the similar profile for both $\psi, \bar\psi$ it is perfectly reasonable that their momentum dependences may have different or even opposite behavior. Besides that the particle/antiparticle effective masses may become different dynamically by a solution

such as those resulting from expressions (6). The hermiticity of Dirac Hamiltonian is not necessarily kept in curved spacetimes due to the time dependence of the metrics that brings time dependence of the spatial position eigenstates (53). Together with the kinetic effects discussed above we just mention topologically non trivial configurations. For the sake of simplicity consider a flat spacetime for which the fermions and antifermions have spatial wavefunctions spherically symmetric with two components: $f(r) = Asech(r/R)h(r)$ and $g(r) = Atanh(r/R)h(r)$, where A is a normalization, $h(r)$ is an overall function that also depends on the other fields, and these two components correspond to the leading terms respectively of the fermion (positive energy) and antifermion (negative energy) solutions. The negative and positive energy solutions could exhibit a topological property similar to the usual scalar field kink, which may guarantee fermions and antifermions to remain separated. Other solutions of the Dirac equation share similar properties (52).

Equations (6,4) do not account for any mechanism of explicit CPT breaking. If CPT breaking (32) took place in the extreme conditions of the early Universe different consequences might be expected, but current observational bounds must be observed in particular CMBR and matter homogeneity. An eventual separation of matter and antimatter was also proposed to occur in different branes which could be close in the higher dimensional space (41).

4.1 Dense systems of antimatter

An usual field theoretical formalism suitable for describing nuclear equation of state in dense stars (56; 83) should also be suitable for dense antimatter physics (24; 49). Similar effective models at the nuclear level are expected to be suitable for describing finite nuclei and hopefully should be suitable for describing the light antinuclei observed recently in relativistic heavy ions collisions of STAR-RHIC and ALICE-CERN (1; 85), although so far one did not have access for the detailed investigation of its structure. These observations just corroborates the expectation of anti-nucleosynthesis in the Early Universe. The energy is not as high as in the early universe GUT scales and baryogenesis is not expected to take place in the same scale. Nevertheless it has been suggested that bubbles of P-violating content of baryons and antibaryons might form in such collisions (5), what might be a preliminar sign for the Sakharov's conditions. However, instead of providing an extensive discussion on these relativistic models for dense baryon equation of state, we present some issues of relevance for finite density anti-baryon systems eventually going beyond usual developments (24). We will refer generally to fermions since the framework is valid, in general lines, for both (anti)baryons or (anti)quarks.

Consider a general finite density environment in Minkowski space in which fermions, either nucleons and eventually hyperons or quarks for quark stars, interact mainly with meson fields, in particular a vector and scalar fields, V_μ, ϕ (49; 83). By considering the Dirac equation for such fermion fields with its eigenvalues and eigenfunctions as functions of the classical counterparts of the meson fields, the energy densities associated to the fermions (anti-fermions) can be calculated straightforwardly $\rho_f[V_0] = Tr\bar{\psi}\hat{H}\psi$. Depending on the truncation of the fermion eigenstates the resulting energy density might exhibit different behaviors. In particular, if the eigenstates of the Dirac equation coupled to the background mean fields (5), it is found that they do not have necessarily the symmetry of the matter-antimatter in the vacuum if V_0 does not change sign. Comparing the behavior for the different densities we notice that antifermion density can be favored for some ranges of V_0, mainly for $V_0 < 0$ as it should be expected (24). It is also interesting that depending on the profile of the solution for $V_0[\rho_B]$, the resulting equation of state can be stiffer or softer.

The resulting expressions for ρ_f might not have the same behavior of those for Fermi liquids (25). Experimental investigation of antinuclei structure will settle further the relevant degrees of freedom and the eventual correspondence of the states of a nucleus and its corresponding anti-nucleus.

As discussed in Section 3, in the same way as quarks can form superconducting high density states, antiquarks might also form color superconducting states with di-antiquark condensation ($< \bar{q}\bar{q} >$) and it could take place in dense antiquark stars (23; 70). Some effects of classical tensor and vector fields, eventually associated to classical gluon configurations (which could be relevant for the time of the quark production), are usually considered for the formation of superconductive states at very high densities (22).

5. Summary

Observational Cosmology is coming to a new era where large observatories start to provide much more precise data. Antimatter asymmetry and states in the Universe have been investigated for short period of time and much has to be done to establish deeper knowledge about them. At the end we remain deeply inside a matter dominated region of the Universe, even if antimatter objects might be not so far. CMBR fluctuations are the most well established observables that induce constraints in the evolution of structures and diffuse gamma ray spectra can provide direct fingerprint of antimatter annihilation and therefore is suitable for investigating anti-objects and their interaction with matter. Cosmic rays are currently measured with high precision and sensitivity, although the corresponding interpretation is not obvious and depends strongly on further models for matter-antimatter interactions in the interstellar medium.

Most observables favor a highly asymmetric $B - \bar{B}$ Universe for which there is a difficulty of establishing the realistic mechanisms which allowed for the baryogenesis according to Sakharov conditions, even if other scenarios have been proposed. In particular the mechanism(s) of CP breaking that really contributed in the Early Universe are not fully elucidated, and several of them cannot be tested in laboratories. If there is hidden antimatter, or islands of antimatter not observed so far, so that the Universe is rather (or nearly) symmetric, then mechanisms that provided the formation of anti-structures (different or not from the known matter structures) or such spatial separation between matter and antimatter must be understood. CPT, gravitational interaction of antimatter eventually with the formation of primordial anti-black holes, among other objects, are some of the issues deserving further theoretical and observational investigation.

6. Acknowledgements

7. References

[1] Abelev, B.I. et al, (2010). *Science* 2, Vol. 328, pp. 58-62. Agakishiev, H. et al, STAR-RHIC Collaboration, arXiv:nucl-ex/1103.3312.

[2] Abdo, A. A. et al, (2009) *Phys. Rev. Lett.* 103, 251101 (2009),

[3] Adamson, P. et al. (MINOS Collaboration), (2010) *Phys. Rev. Lett.* 105, 151601. Engelhardt, N. , Nelson, A. E., Walsh, J. R., arXiv:hep-ph/1002.4452.

[4] AEGIS Collaboration, URL: http://aegis.web.cern.ch/aegis

[5] Aggarwal, M. M., et al. (STAR Collaboration) *Phys. Rev. Lett.* 106 (2011) 62002.

[6] Madsen N. (2010) Philosophical Transactions of the Royal Society A 368, vol 1924, 3671.

[7] Alcock , et al., (1997) *Astrophys. J.* 486, 697. MACHO project, URL http://www.macho.mcmaster.ca/

[8] Arkani-Hamed. N, Dimopoulos. S, Dvali. G, (1998) *Phys. Lett.* B 429 (1998) 263; Antoniadis. I, et al, *Phys. Lett.* B 436 (1998) 257. Dolgov.A. D, Urban.F.R, (2006) *Nucl. Phys.* B752 , 297; hep-ph/0605263; Bambi.C, Dolgov.A. D, Freese. K, (2007) *Journal of Cosmology and Astroparticle Physics* 04, 005; hep-ph/0612018.

[9] Affleck. I, Dine. M, (1985) *Nucl. Phys.*, B249 , 361.

[10] AMS Collaboration, *arXiv:hep-ex/0002048v1*; URL: http://ams.cern.ch.

[11] AMS-2, URL: http://ams-02project.jsc.nasa.gov .

[12] ASACUSA, URL: http://asacusa.web.cern.ch/ASACUSA/

[13] ATRAP, URL: http://public.web.cern.ch/public/en/research/ATRAP-en.html

[14] Baltz, E. and Edsjo J., (1999) *Phys. Rev.* D 59, 023511.

[15] Bambi. C, Dolgov. A. D, (2007) *Nucl. Phys.* B 784, 132, astro-ph/0702350.

[16] Bambi. C, et al, (2009) *JCAP* 9, 013. Treves. A., Turolla. R,(1999) *Astrophys. J.* 517 , 396.

[17] Belotsky. K. M, Golubkov.Yu.A, Khlopov. M.Yu,. Konoplich. R.V., Sakharov. A.S, (2000) *Phys. Atom. Nucl.* 63, 233; Golubkov. Y. A, Khlopov. M. Y, (2001) *Phys. Atom. Nucl.* 64, 1821; astro-ph/0005419; Fargion. D, Khlopov. M, (2003)*Astropart. Phys.* 19, 441, hep-ph/0109133; Casadei D., *astro-ph/*0405417, revised version in 2008.

[18] Bergstrom. L, Edsjo. J and Ullio. P, (1999) *Phys. Rev.* D 59, 43506. G. Jungmamn and M. Kamionkowski, (1994) *Phys. Rev.* D, 49 , 2316.

[19] BESS Collaboration URL: http://www.universe.nasa.gov/astroparticles/programs/bess/

[20] See for example: Bimonte G., Lozano G., (1996) *Phys. Lett.* B 366, 248.

[21] Birrell. N. D and Davies. P. C. W, (1982) *Quantum fields in curved space*, Cambridge Univ. Press, Cambridge. Fulling. S. A, (1996) *Aspects of Quantum Field Theory in Curved Space-Time*, London Math Society, S.T. 17, C. U. P., Cambridge.

[22] See for example Blaschke, D et al., (2005) *Phys. Rev.* D 72, 065020.

[23] Braghin, F.L.,(2007) *Nucl. Phys.* A 790, 546c, hep-ph/0611390.

[24] Braghin, F.L. (2007) *Int. Journal of Mod. Phys.* D 16, 96.

[25] Braghin, F.L. (2010) *Int. Journal of Mod. Phys.* D 19, 1505.

[26] Brown. R. W and Stecker. F. W, (1979)*Phys. Rev. Lett.*, 43 , 315.

[27] Burles, S., Nollett, K. M., Turner, M. S., (2001) *The Astrophysical Journal*, 552.

[28] Casadei,D., astro-ph/0405417. Golubkov,Y.A., Khlopov, M.Y., (2001) *Phys. Atom. Nucl.* 64, 1821; astro-ph/0005419.

[29] CBI: URL: http://www-astro.physics.ox.ac.uk/research/expcosmology/groupcbi.html

[30] Cohen. A. G, Kaplan. D. B, .Nucl. Phys., B308 (1988) 913.

[31] Cohen A. G, de Rújula. A, Glashow. S. L, (1998) *The Astrophys. Journ.* 495, 539 . Auriemma. G, (2003) *Chin. J. Astron. Astrophys.* 3, 30.

[32] See for example: Kostelecky, A., Russell, N., (2008) Data Tables for Lorentz and CPT Violation, hep-ph, arXiv:0801.0287v3 Mavromatos, N. E., (2005) hep-ph/0504143. Russell, N., (2005) hep-ph/0511262.

[33] De Simone A, et al, (2010) *Journ. of Cosmology and Astroparticle Phys.* 7, 17.

[34] Dine M., Kusenko. A, (2003) *Rev. of Mod. Phys.* 76 1 and references therein.

[35] Dolgov. A. D, (1992) Phys. Repts 222, No. 6.

[36] Dolgov. A. D, (2005) *Lectures presented at Varenna* Varenna, Italy, hep-ph/0511213.

[37] Dolgov,A.D., Kawasaki, M., Kevlishvili, N. (2008) hep-ph/arXiv:0806.2986v2

[38] Dolgov, A.D. (2009) arXiv:0901.2100v1.

[39] Dolgov,A.D., (2011) *Phys. of At. Nuclei* 74, 3, 462; hep-ph/1002.2940v1

[40] Doser, M. (AEGIS collaboration) (2010) *Journ. of Physics: Conf. Series* 199, 012009.

[41] Dvali. G. R, Gabadadze. G, Porrati. M, (2000) *Phys. Lett.* B485, 208, hep-th/0005016.

[42] FAIR/GSI facility: URL: https://www.gsi.de/portrait/fair$_e$.html

[43] FERMI Collaboration: URL: http://fermi.gsfc.nasa.gov/.

[44] Fixen. D.J et al., (1996) *Astrophys. J.* 473, 576. Mather . C.J et al., (1994) *Astrophys. J.* 420, 439. Bersanelli. M et al., (1994) *Astrophys. J.* 424, 517.

[45] Fukugita. M, Yanagita. T, (1986) *Phys. Lett.*, B174, 45. Buchmuller. W, Di Bari. P, Plumacher. M, (2004) *New J. Phys.*, 6 , 105.

[46] Gast H. et al, (2008) *Proceedings of the 30th International Cosmic Ray Conference*, ed by Caballero. R. et al Mexico City, Mexico, p. 293.

[47] Georgi. H ,Glashow. S L , (1974) *Phys. Rev. Lett.* 32, 438.

[48] Goudelis. A, Mambrini. Y and Yaguna. C, (2009) *JCAP*12 ,008

[49] Greiner, W., (2003) *AIP Conf. Proc.* 1323, 109; *Physics of Atomic Nuclei*, 66, 1009.

[50] Hajdukovc, D.S. , (2010) *Astrophys. and Space Science* 330, 1.

[51] Hu. P. H, Sarkar. U, (2010) *Phys. Rev. D* 81, 033001.

[52] Hirayama,M., Ninagawa,S. (1989) *Phys. Rev.* D 39, 1602 . A different approach in: Góźdź,M., Nakonieczny, L., Rogatko,M.(2007) *Phys. Rev.* D 81, 104027.

[53] Huang,X., Parker,L. (2009) *Phys. Rev.* D79, 024020.

[54] Several works for example in *Int. J. Mod. Phys.* D 17 (2002); and in Ref. (55). Santilli,R.M. , (1998) *Int. Journ. of Mod. Phys.* D 7, 351.

[55] Itzykson, C. and Zuber, J.-B. (1985) *Quantum Field Theory*, McGraw Hill, Singapore. Conditions for CPT violation, for example, in: Mavromatos, N. E., hep-ph/0504143. Tests on CPT, for example: N. Russell, hep-ph/0511262.

[56] Several contributions for the Proceedings of International Workshop on Relativistic Astrophysics, (2010) Int. Journ. of Mod. Phys. D 19. Ed. by J.E. Horvath, M.M.B.M. de Oliveira, B.E.J. Bodmann, C.A.Z. Vasconcellos, H. Stoecker, W. Greiner.

[57] Kinney, W.H. et al (1997) *Phys. Rev. Lett.* 79. A. G. Cohen, A. De Rujula, astro-ph/9709132.

[58] Kogut, J. et al (1983) *Phys. Rev. Lett.* 50, 393 Berges, J., Tetradis N., Wetterich C., (2002) *Phys. Rep.* 363, 223.

[59] Kohri,K., Mazumdar, A., Sahu, N., (2009) *Phys. Rev.* D 80, 103504.

[60] Kuzmin. V. A, Shaposhnikov. M. E, Tkachev. I. I, (1981) *Phys. Lett.* B105, 167; Kuzmin. V. A, Tkachev.I. I, Shaposhnikov. M. E, (1981) *Pisma Zh. Eksp. Teor. Fiz.* 33, 557.

[61] Langlois,D., C. R. (2003) *Physique* 4, 953.

[62] Lee,T.D. (1973) *Phys. Rev.*, D8 , 1226. Brown, R.W., Stecker, F.W. (1979) *Phys. Rev. Lett.* 43, 315.

[63] Lee, T.D., (1974) *Phys. Rept.* 9, 43. Brown, R.W. and Stecker, F.W. (1979) *Phys. Rev. Lett.* 43 315.

[64] Maki. K, Mitsui. T, T., and Orito. S, (1996) *Phys. Rev. Lett.* 76, 3474. Hawking,. S. W, (1971) *Monthly Notices Roy. Astron. Soc*, 152, 75. Kiraly. P et al., (1981) *Nature*, 293, 120.

[65] Mori. K, et al.,(2002) *Astrophys. Journal* 566 , 604 See e.g.:URL: http://www.cita.utoronto.ca/ kaya/gaps.html

[66] Morselli,A., Moskalenko, I.V., *Status of indirect searches in the PAMELA and FERMI era Proceedings of Science, PoS*, http://pos.issa.it (IDM2008) 025; arXiv:0811.3526

[67] Moskalenko. I. V., et al (2002) *Astrophys. J.* 565, 280. Tan. L. C and Ng. L. K , (1983) *J. Phys.* G 9, 227; (1983) *J. Phys.* G 9 , 1289.

[68] Naselsky P.D.,Chiang, L-Y (2004) *Phys. Rev.* D69, 123518.

[69] For example in Nishino, H. , et al. (Super-K Collaboration) (2009). *Phys. Rev. Lett.* 102, 141801.

[70] Oaknin. D. H, Zhitnitsky. A, (2005) *Phys. Rev.* D71, 023519, hep-ph/0309086; Oaknin. D. H, Zhitnitsky. A, (2005) *Phys. Rev. Lett.* 94 , 101301, hep-ph/0406146. K. Lawson and A.R. Zhitnitsky, arXiv:0704.3064

[71] Orito. S et al. (BESS Collaboration), (2000) *Phys. Rev. Lett.* 84, 1078; astro-ph/9906426.

[72] PAMELA Collaboration, URL: http://pamela.roma2.infn.it/index.php.

[73] Parker L. E, Toms. D. J, (2009) *Quantum Field Theory in Curved Spacetime*, Cambridge Univers. Press, Cambridge.

[74] Picozza. P et al., *Astroparticle Physics* 27 (2007) 296. Boezio M. et al, PAMELA Collaboration, (2008). *Journal of Physics: Conference Series*, 110, 062002.

[75] Picozza. P, Morselli. A, (2008) *Journal of Physics: Conference Series* 120,042004, IOP Publishing.

[76] Purcell. W. R et al., (1997) *Astrophys. J.* 491 , 725. Milne. P. A, et al (2002) *New Astron. Rev.* 46 , 553. Knodlseder. J. et al., (2005)*Astron. Astrophys.* 441, 513. Jean. P. et al., (2006) *Astron. Astrophys.* 445, 579.

[77] Raffelt, G. G., (2007) arXiv:astro-ph/0701677.

[78] Riotto, A., Trodden, M. (1999) Annu. Rev. Nucl. Part. Sci. 49:35.

[79] Sakharov. A.D, (1967) *JETP Lett.* 6, 24.

[80] Sannino. F and Schafer. W hep-ph/0111098.

[81] See for example Sarkar, U. (2000) *PRAMANA* Indian Academy of Sciences 54, 101.

[82] Sasaki. M *et al*, (2008) *Advances in Space Research* 42 , 450ï£¡454.

[83] Serot. B. D and Walecka. J. D, (1997) *Int. Jour. Mod. Phys.* E 6, 515 and references therein.

[84] Shaposhnikov. M, (2009) *Progr. of Theor. Phys.*122, 185.

[85] Sharma, N., ALICE Collaboration, arXiv: nucl-ex/1104.3311v2.

[86] Sherk. J, (1979) *Phys. Lett.* B 88 , 265.

[87] Simon. M, Molnar. A and Roesler.S, (1998) *Astrophys. Journal* 499 , 250. Bergstrom.L , et al, (1999) *Astrophys. Journal* 526 , 215

[88] See for example Singh. S, (2004) *Big Bang: The Origin of the Universe*, HarperCollins,.

[89] Spada. F., *hep-ex*/0810.3831v1.

[90] Spergel. D.N. et al., (2003) *Astrophys. J.* Suppl. 148, 175; astro-ph/0302209.

[91] Sreekumar, P. et al., (1998) *Astrophys. Journal* 494 , 523.

[92] Steigman. G, (1976) *Ann. Rev. Astron. Astrophys.* 14 , 339.

[93] Steigman G., (2008) *J. Cosmol. Astropart. Phys.* JCAP 10, 001. G. Steigman, arXiv:0808.1122.

[94] Streater, R. F., Wightman,A. S. (1964) *PCT, spin and statistics, and all that*. Benjamin/Cummings.

[95] Strong, A. W., Moskalenko, I. V., and Ptuskin, V. S., (2007) *Annu. Rev. Nucl. Part. Sci.* 57, 285.

[96] t'Hooft. G, (1976) *Phys. Rev. Lett.* 37 8; (1976) *Phys. Rev.* D14 , 3432.

[97] Izotov, Y. I., Thuan, T. X., (1998) *Astrophys. J.* 500, 188.

[98] Yao et al, (2006) *Journal of Physics* G 33 , 1.

[99] Yokoyama J., et al, (1987) *Int. J. Mod. Phys.* A 2 , 1808; J. Yokoyama, et al (1988) Prog. Theor. Phys. 79 , 800. D. Grigoriev, hep-ph/0006115.

[100] Weidenspointner . G. et al., (2006) *Astron. Astrophys.* 450, 1013 ; astro-ph/0601673.

[101] Weidenspointner.G, et al., *astro-ph*/0702621.

[102] Weinberg, S., (1975) *Gravitation and Cosmology*, John Wiley & Sons, New York.

[103] WMAP Collaboration, URL: http://map.gsfc.nasa.gov/mission/ E. Komatsu, (2011) *Astrophys.J.Suppl.*192, 18; arXiv:astro-ph/1001.4538.

[104] See for example: Wong, C.-Y., Lee, T.G., arXiv:physics/1103.5774.

[105] Zeldovich, Ya.B., Kobzarev, I.Yu. Okun, L.B. , (1974) *Zh. Eksp. Teor. Fiz.* 67, 3; *Sov. Phys. JETP* 40 (1974) 1.

[106] Zeldovich Ya. B, (1976) *Pisma Zh. Eksp. Teor. Fiz.* 24, 29; Dolgov.A. D. (1981) *Phys. Rev. D* 24,1042.

[107] Zeldovich. Ya.B., Kobzarev. I.Yu, Okun. L. B, (1974) *Zh. Eksp. Teor. Fiz.*, 67, 3.

[108] Zhang, J., Yuan,Q., Bi, X.-J. (2010) *The Astrophysical Journal*, 720, 9.

Galaxy Rotation Curves in the Context of ΛCDM Cosmology

Marc S. Seigar[1] and Joel Berrier[2]
[1]*University of Arkansas at Little Rock*
[2]*University of Arkansas at Fayetteville*
U.S.A.

1. Introduction

The motion of stars and gas in spiral galaxies provide a means of measuring the density profile of such galaxies. In the late 1960s, Vera Rubin started to take observations of the rotation velocities of disk galaxies (otherwise known as rotation curves). One of the most famous rotation curves she obtained was that of M31 (Rubin & Ford 1970). A key result of this study is that the rotation velocity of M31 remains high at very large radii ($r > 20$ kpc; see Fig. 1). The conclusion from these rotation curve data is that galaxies must contain much more mass than the visible light would otherwise indicate.

These so-called flat rotation curves were confirmed in a series of papers in the late-1970s (Rubin et al., 1977, 1978a, b, 1979; Peterson et al., 1978) and early 1980s (Rubin et al., 1980, 1982, 1985). These rotation curves have now been extended to much larger radii with sensitive HI measurements (e.g., Carignan et al., 2006) demonstrating that the rotation velocities remain high far beyond the optical disk in galaxies. This suggests that the visible light (and thus stellar mass) accounts for only a small fraction ($\sim 15\%$) of the mass in spiral galaxies. This still remains one of the best pieces of evidence in favor of dark matter cosmology.

The outline of this chapter is as follows: Section 2 describes the growth of structure in the Universe and how halos of cold dark matter assemble over cosmological times. Section 3 describes one of the fundamental problems with cold dark matter cosmology, that simulations predict cuspy central densities in dark matter halos, yet observations (particularly of dwarf galaxies) prefer constant density cores. Section 4 describes the Tully-Fisher zeropoint problem, the fact that cosmological models cannot reproduce the luminosity-rotation velocity relation for galaxies (Tully & Fisher, 1977) without overproducing the number density of galaxies at fixed luminosity. Finally, in Section 5 we make some conclusions about the present state of cosmological simulations on galaxy-sized scales.

2. Dark matter halo assembly and the growth of structure

The basic assumptions of modern cosmology are that the Universe is homogeneous and isotropic (LeMaitre, 1931). While these assumptions work on large scales it is clear that they are violated on cosmologically small scales (i.e., galaxy-sized scales). The structure of the Universe becomes much more complicated at these scales, with the growth of structure dominated by gravity. The standard model for the cosmology of the Universe has proven to
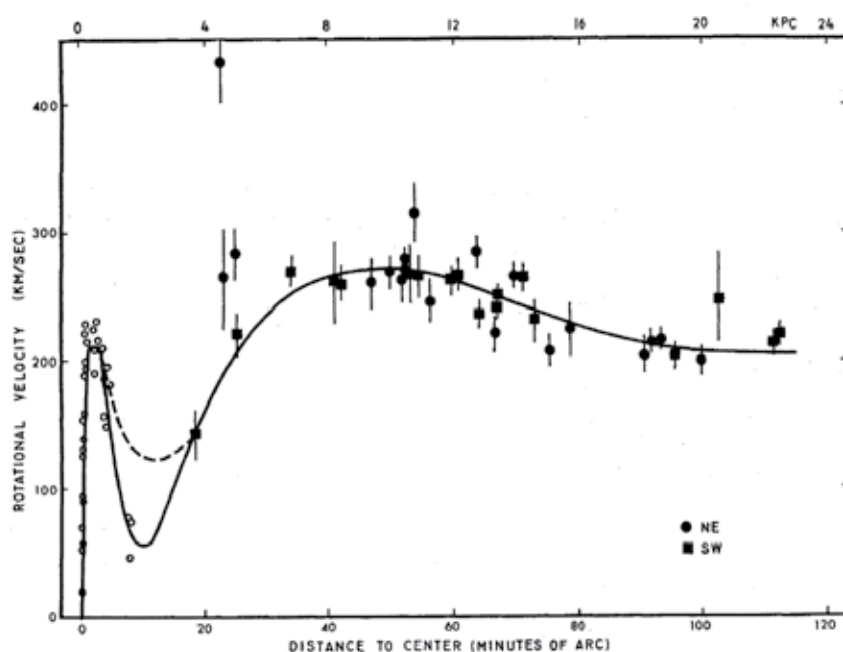
Fig. 1. Rotation curve of M31 from Rubin & Ford, (1970). Rotational velocities for OB associations in M31, as a function of distance from the center. The solid curve is the adopted rotation curve. For $r < 12'$, the curve is a fifth-order polynomial; for $r > 12'$ the curve is a fourth-order polynomial which is required to make the rotation curve approximately flat near $r = 120'$.

be an invaluable tool as this cosmology reproduces the large scale structures observed in the Universe.

The initially small quantum perturbations in the early Universe grow rapidly due to inflation in the standard ΛCDM cosmology. These overdense regions provide the seeds for gravitational collapse. These collapsing overdensities, which are primarily composed of dark matter, provide the initial potential wells for baryons to condense and begin the process of galaxy formation. Observations of the cosmic microwave background (CMB; Spergel et al., 2007) show that the initial perturbations appear to have a Gaussian random distribution. The idea is that these initial perturbations grow to become the galaxies and clusters of galaxies that we see in the local Universe.

The fluctuations in the Universe are characterized by a power spectrum which may be expressed as the Fourier transform of the two-point correlation function — the average number of pairs in a spherical shell at radius r of thickness dr above what is expected from a purely random distribution (Peebles, 1980). Thus, the correlation function measures the clustering of galaxies. Another means to express these perturbations is as the mass variance of the fluctuations. Galaxy surveys are able to calculate the mass density fluctuations within a spherical region to measure the rms variance of the Universe. Early surveys suggested that a sphere of radius 8 Mpc was needed to bring the variance to unity. This standard measurement was demonstrated to be incorrect, but it is still traditional to use σ_8 as a measure of the clumpyness of matter in the Universe. Current cosmologies from WMAP results combined with other data sets place $\sigma_8 = 0.816 \pm 0.024$ (Komatsu et al., 2011).

The overdensities are the seeds for structures to grow through gravitational interactions. Observations provide abundant evidence that the structure in the Universe formed hierarchically (Blumenthal et al., 1984; Lacey & Cole, 1993). In the regions which are overdense compared to the mean density of the Universe, the first dark matter halos begin to form after inflation. Dark matter dominates the total matter density of the Universe, and, as it does not interact with radiation, is the first matter to undergo spherical collapse due to gravity.

The initial power spectrum of the perturbations produces few large potential wells. Early halos of dark matter will grow through two processes. The first of these is smooth accretion of additional material. The second growth mechanism is mergers with other dark matter halos. In the standard cosmological paradigm the merger rate of distinct dark matter halos is robustly predicted. Several studies have predicted that the rate of major mergers for galaxy-sized cold dark matter halos scales with redshift as $(1 + z)^m$, where the exponent lies in the range $2.5 \leq m \leq 3.5$ (e.g. Governato et al., 1999; Gottlober et al., 2001).

We expect that galaxies are formed inside larger dark matter halos (White & Rees, 1978; White & Frenk, 1991). The early galaxies begin to form as the potential wells created by the dark matter halos draw baryonic matter into the halo, allowing galaxies to form earlier than would otherwise be possible. These galaxies grow through mass accretion as additional material is drawn into the deepening potential well of the dark matter halo. Galaxy mergers allow the growth of more massive galaxies as the Universe continues to age. These merger events are key in altering galaxy morphologies by growing bulges in disk galaxies, transforming disks into ellipticals and inducing star formation.

Comparisons between predictions from models of a ΛCDM Universe and observations show remarkable agreement on large scales (Blumenthal et al., 1984; Springel et al., 2005). The two point correlation function of galaxies is similar to the measured two point function for dark matter halos. Simulations of large volumes of the Universe in this cosmology reproduce the clustering and distribution of galaxies on large scales.

3. Spiral galaxy rotation curves and the cusp-core problem

Despite the successes of the ΛCDM cosmology at describing the large scale structure of the Universe, the model remains far from perfect. Observations show significant differences on small-scales from the theoretically predicted structure of ΛCDM. A detail of significant difference comes in the density profiles of dark matter halos. N-body simulations of structure growth have shown that galaxy-sized halos have density profiles with central cusps (e.g., Navarro et al., 1997), yet observations of late-type galaxies tend to favor a constant density core (e.g., Kuzio de Naray, 2008). In the mid 1990s, some authors realised that cuspy profiles would prove difficult for a CDM cosmology (e.g., Flores & Primack, 1994; Moore, 1994)

It has been 40 years since the discovery that galaxies are surrounded by extended massive halos of dark matter (e.g., Rubin & Ford, 1970). A variety of observational probes—disk rotation curves, stellar kinematics, gas rings, motions of globular clusters, planetary nebulae and satellite galaxies, hot gaseous atmospheres, and gravitational lensing effects—are now making it possible to map halo mass distributions in some detail. These distributions are intimately linked to the nature of the dark matter, the way halos formed, and the cosmological context of halo formation.

Insight into these links came first from analytic studies. Building on the early work of Gunn & Gott, (1972), similarity solutions were obtained by Fillmore & Goldreich, (1984) and Bertschinger, (1985) for the self-similar collapse of spherical perturbations in an Einstein–de Sitter Universe. Such solutions necessarily resemble power laws in the virialized regions. Hoffman & Shaham, (1985) and Hoffman, (1988) extended this analysis by considering open

universes and by modeling as scale-free spherical perturbations the objects that form by hierarchical clustering from power-law initial density perturbation spectra [$P(k) \propto k^n$]. They argued that isothermal structure ($\rho \propto r^{-2}$) should be expected in an Einstein–de Sitter Universe if $n \leq -2$ and that steeper profiles should be expected for larger n and in open universes.

Despite the schematic nature of these arguments, their general predictions were verified as numerical data became available from N-body simulations of hierarchical cosmologies. Power-law fits to halo density profiles in a variety of simulations all show a clear steepening as the density of the Universe decreases (Frenk et al., 1985, 1988; Quinn, Salmon & Zurek, 1986; Efstathiou et al., 1988; Zurek, Quinn & Salmon, 1988; Warren et al., 1992; Crone, Evrard & Richstone, 1994). An apparent exception was the work of West, Dekel & Oemler, (1987), who found that *galaxy cluster* density profiles show no clear dependence on n.

Significant departures from power-law behavior were first reported by Frenk et al., (1988), who noted that halo profiles in cold dark matter (CDM) simulations steepen progressively with increasing radius. Efstathiou et al., (1988) found similar departures—at odds with the analytic predictions—in their simulations of scale-free hierarchical clustering. They also noted that these departures were most obvious in their best resolved halos. Similar effects were noted by Dubinski & Carlberg, (1991) in a high-resolution simulation of a galaxy-sized CDM halo. These authors found their halo to be well described by a density profile with a gently changing logarithmic slope, specifically the one proposed by Hernquist, (1990).

In a series of papers by Navarro et al., (1995, 1996, 1997), these authors used high-resolution simulations to study the formation of CDM halos with masses spanning about 4 orders of magnitude, ranging from dwarf galaxy halos to those of rich galaxy clusters. They showed that the equilibrium density profiles of CDM halos of all masses can be accurately fitted over two decades in radius by the simple formula:

$$\frac{\rho(r)}{\rho_{\text{crit}}} = \frac{\delta_c}{(r/r_s)(1 + r/r_s)^2},\tag{1}$$

where r_s is a scale radius, δ_c is a characteristic (dimensionless) density, and $\rho_{\text{crit}} = 3H^2/8\pi G$ is the critical density for closure. This profile, dubbed the NFW profile, differs from the Hernquist (1990) model only in its asymptotic behavior at $r >> r_s$ (it tends to r^{-3} instead of R^{-4}). Power-law fits over a restricted radial range have slopes that depend on the range fitted, steepening from -1 near the center to -3 at large r/r_s.

While the predictions of these simulations are self consistent they differ from the observed rotation curves of late-type disk and dwarf galaxies. These quite often show mass distributions with lower than predicted densities or with constant density cores, where $\gamma \simeq 0$. These observations would suggest a preference for a pseudo-isothermal density profile (Gentile et al., 2004, 2005; Kuzio de Naray et al., 2006, 2008; Shankar et al., 2006; Simon et al., 2005; Spano et al., 2008; Swaters et al., 2003; see Fig. 2). This problem is referred to as the cusp-core problem due to the significant differences between the observed measured dark matter halo density profile and the halo profiles derived from theory and simulations.

Studies that reveal constant density cores in galaxies have concentrated on dwarf galaxies and low-surface brightness galaxies. Swaters et al., (2003) obtained high-resolution Hα rotation curves for 15 dwarf and low surface brightness galaxies. They found that the central mass distributions of these galaxies follow a power law at small radii, $\rho(r) \propto r^{-\alpha}$, with inner slopes in the range $0 < \alpha < 1$ for most of their galaxies, and claim that the inner slopes are therefore poorly constrained. In general, they claim that galaxies with constant density cores (i.e., $\alpha = 0$)
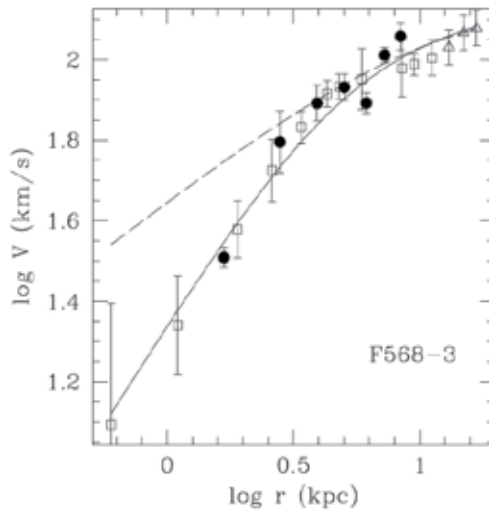
Fig. 2. Rotation curve from Kuzio de Naray et al., (2006) for the galaxy F568-3. A pseudo-isothermal fit (solid line) and a forced NFW fit (dashed line) to rotation curve IFU data (black points) is shown. This is combined with long-slit data (open squares) from de Blok et al., (2001) and HI data (open triangles) from de Blok et al., (1996). The fit is in the limit of no baryons, and clearly shows that the data is better described by a cored halo (or pseudo-isothermal fit) than a cuspy halo (or NFW fit), particularly at small radii.

provide somewhat better fits, but the majority of their sample (around 75%) are also consistent with $\alpha = 1$. An NFW-type model (with $\alpha = 1.5$) was however ruled out in every case.

A potential solution for this conflict could be that the observational results are affected by a significant amount of non-circular motion in the central regions of a galaxy. Triaxial halos (or asphericity) or a strong bar in the central regions of a halo have been suggested as a means to include significant non-circular motion in the central regions of galaxies (Hayashi et al., 2007; Bailin et al., 2007). These non-circular motions may result in observations of lower than expected velocities in the central regions of galaxies. However, several studies have shown that, even in the presence of a triaxial halo, and taking into account the non-circular motions induced by such a halo, it is difficult to reproduce the dark halo densities expected by ΛCDM (e.g., Gentile et al, 2005).

Low surface brightness galaxies and dwarf galaxies are dominated by dark matter content at all radii. As a result, they provide an ideal test case in which baryonic physics cannot affect the shape of the dark matter halo as strongly. Most studies, to date, have concentrated on these late-type galaxies, and below we describe the most significant contributions to the field. Gentile et al., (2004, 2005) studied a total of 6 galaxies (including DD0 47 in their 2005 paper) using a combination of HI and Hα rotation data. They decompose their rotation curve data into stellar, gaseous and dark matter contributions, and their inferred density distribution was compared with various model mass distributions. Their observations point to a power law density distribution that follows either $\rho(r) \propto r^{-2/3}$ or a constant density core, i.e., $\rho(r) \propto r^0$, where the core size $r_{\text{core}} \simeq r_{\text{opt}}$. They conclude that ΛCDM models are in clear conflict with their data, even in the case of DDO 47, which has a clear triaxial halo.

Simon et al., (2005) used a combination of high-resolution Hα velocity fields and CO velocity fields of five nearby, low-mass galaxies. They find power law density distributions with a power-law index in the range $0.0 < \alpha < 1.2$. They also find that the scatter in α is 0.44, which is three times as large as in ΛCDM simulations. The mean density profile shape that they find is $\alpha = 0.73$, shallower than that predicted by ΛCDM by a factor of ~ 2. Only one galaxy in their sample, NGC 5963, has a cuspy density profile that closely resembles those seen in ΛCDM simulations (see Fig. 3).
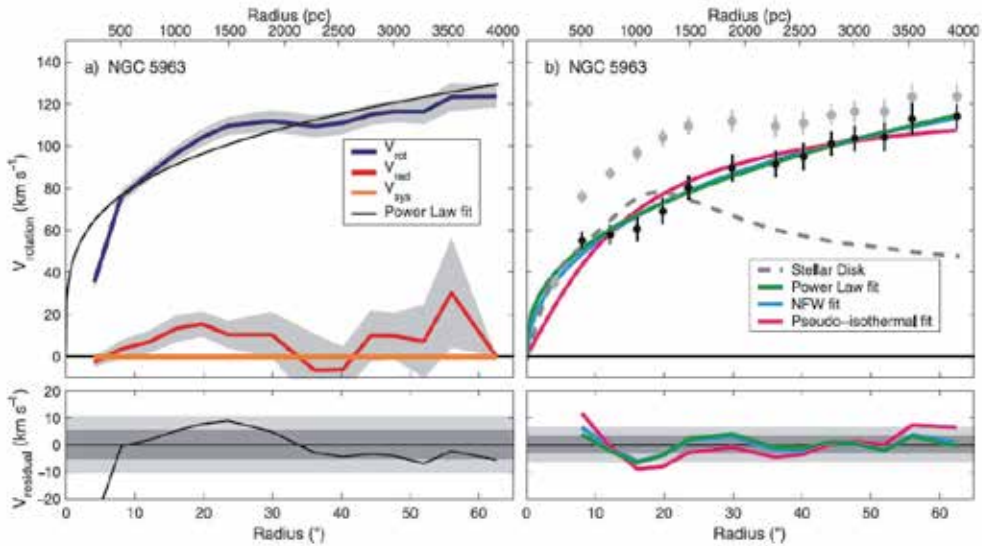

Fig. 3. Rotation curve from Simon et al., (2005) for the galaxy NGC 5963. (a) Tilted-ring model for NGC 5963. The thick blue curve represents the rotation curve. The thick red and orange curves show the radial velocity curve and the systemic velocities, respectively. The shaded grey regions surrounding the rotation and radial velocity curves represent the combined systematic and statistical 1-σ uncertainties. The thin black curve is a power-law fit to the rotation curve. The residuals to this fit are plotted in the bottom panel and the 1- and 2-σ scatter of the data points around the fit is shown by the shaded grey areas. (b) Disk-subtracted rotation curve of NGC 5963 (for $M_*/L_R = 1.24 \, \mathrm{M_\odot/L_{\odot,R}}$). The grey points are the original rotation velocities from panel (a), and the black points are the dark matter halo rotation curve after subtracting the stellar disk ($M_*/L_R = 1.01 \, \mathrm{M_\odot/L_{\odot,R}}$), which is shown as a grey-dashed line. The thick green, cyan, and magenta curves show power-law, NFW, and pseudo-isothermal fits to the halo rotation curve, respectively. The residuals from these fits are displayed in the bottom panel. The 1- and 2-σ scatter of the data points around the power-law fit is shown by the shaded grey areas.

Kuzio de Naray et al., (2006, 2008) observed Hα velocity fields of a sample of 17 low-surface brightness galaxies with the DensePak Integral Field Unit (IFU) on the 3.5-m WIYN telescope at the Kitt Peak National Observatory. In the limit of no baryons, they fit the NFW and pseudoisothermal halo models to the data and found that the rotation curve shapes and halo central densities were better described by the isothermal halo.

One of the most comprehensive studies so far is that of Spano et al., (2006), who presented high-resolution *Hα* two-dimensional velocity fields of 36 spiral galaxies. They combined

their kinematical data with photometric data and derived mass models using two different functional forms: an isothermal sphere and an NFW profile. Their results point to the existence of a constant density core in the centers of dark matter halos (i.e., a power law slope of $\alpha = 0$) rather than a cuspy core for all Hubble types from Sab to Im.

In an effort to investigate whether it is possible to confuse a cuspy-type central density profile for a constant density core due to the observational setup, Kuzio de Naray & Kaufmann, (2011) performed a study of simulated LSB galaxies in simulations with mock observations using the parameters of the DensePak IFU. Their simulations included cuspy, cored, and triaxial dark matter halos. Using mock observations of the simulated galaxies under a variety of conditions, Kuzio de Naray & Kaufmann, (2011) were able to identify signatures in the velocity fields which allowed the recovery of the correct halo information. From this study, and that presented by Gentile et al., (2005), it is natural to conclude that these late-type, low surface brightness and dwarf galaxies really do have central density profiles that are inconsistent with the expectations of ΛCDM cosmology. The question now is, how can we explain these inconsistencies?

In contrast to the results presented so far, one of the best-studied and closest galaxies, M33, shows clear evidence of a central density cusp. This is consistent with the expectations of the NFW profile and ΛCDM cosmology (Corbelli & Salucci 2000; Seigar 2011; see Fig. 4). Given the late Hubble type (Scd; de Vaucouleurs et al., 1991) and the small size of M33, one might expect it to have a density profile consistent with those of the late-type galaxies described above. Nevertheless, the two studies of the halo density profile of M33 from the last decade have shown that it does seem to conform to the expectations of ΛCDM cosmology. While Corbelli & Salucci (2000) find a dark halo mass of $M_{\mathrm{vir}} = 7.4 \times 10^{10}$ M$_\odot$ and a halo concentration of $c_{\mathrm{vir}} = 5.6$, and Seigar (2011) finds a dark halo mass of $M_{\mathrm{vir}} = 2.2 \times 10^{10}$ M$_\odot$ and a halo concentration of $c_{\mathrm{vir}} = 4.0$, both studies conclude that an NFW profile clearly provides a better fit to the data than a pseudo-isothermal profile. However, given the discrepancy in the dark halo virial mass between these two studies, it is worth noting the differences in the analyses. The difference really comes down to how the stellar mass distribution is derived. To derive the stellar mass, Corbelli & Salucci, (2000), use a rather old K band image from Regan & Vogel, (1994). Seigar (2011) uses a Spitzer 3.6-μm image, which has a far superior signal-to-noise ratio. Furthermore, Corbelli & Salucci (2000) assume a distance to M33 of 0.7 Mpc, whereas Seigar (2011) adopts the more accurate measurement of 0.84 Mpc (Magrini et al., 2007). As a result of this underestimate of the distance to M33, Corbelli & Salucci (2000) have underestimated the size of the visible galaxy by ∼17%, and this will propagate into a much larger error in their total mass estimate for the dark halo. It should also be noted that the dark halo mass estimate provided by Corbelli & Salucci (2000) is only around 25% less than the best estimates for M31 of $M_{\mathrm{vir}} \simeq 10^{12}$ M$_\odot$ (e.g., Seigar et al., 2008). In terms of visible size, M33 is about 20% the size of M31, and, if we assumed that the halo mass scaled (more or less) linearly with visible size, one would naturally expect a halo mass similar to that found by Seigar (2011).

Despite the work on M33 pointing towards a dark halo density profile that is consistent with ΛCDM cosmology, the many other studies (already mentioned above) clearly indicate that there is a problem. Several possible solutions to resolving this conflict have been suggested. One possibility is that these observations are pointing to a real problem with Cold Dark Matter cosmology, perhaps indicating that the dark matter is not cold, but rather warm (Zentner & Bullock, 2002), or possibly even a form of self interacting dark matter (e.g. Spergel & Steinhardt, 2000; Kaplinghat et al., 2000), in which case it is easier to produce constant density cores at the centers of dark matter halos. More recent results from Kuzio de Naray et al.,
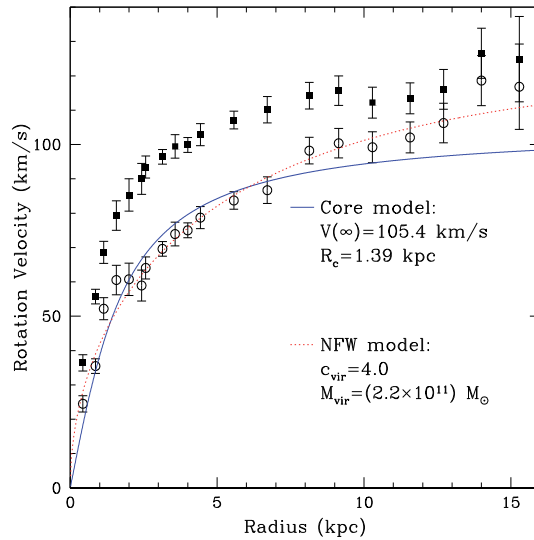
Fig. 4. Rotation curve from Seigar, (2011) for the galaxy M33. This figure shows the HɪI rotation curve from Corbelli & Salucci, (2000) modeled using a pseudo-isothermal model (core model; blue-solid line) and a NFW model (red-dotted line). The squares represent the total rotation velocities, and the circles represent the contribution of the dark matter to the rotation velocities (after subtraction of the stellar and gas mass distributions).

(2010) suggest these scenarios are unlikely due to discrepancies in the size of the cores created in these models and phase space densities that are orders of magnitude too small. In this paper, the authors present $H\alpha$ velocity fields of 9 low-surface brightness galaxies, and they fit the data with models representing warm dark matter and self-interacting dark matter. They find that the inferred core radii increase with halo mass (whereas warm dark matter predicts that the core radius should decrease with halo mass). They also find that the dark matter halo core densities vary by a factor of ∼30 from galaxy to galaxy, while showing no systematic trend with the maximum rotation velocity of the galaxy. Their result strongly argues against the core size being directly set up by large self-interactions (whether scattering or annihilation) of dark matter. One would therefore naturally conclude that the inferred cores do not provide motivation to prefer warm or self-interacting dark matter over cold dark matter.

Another possible explanation for the prevalence of constant density cores in these late-types galaxies is because they form late in the history of the Universe (Wechsler et al., 2002), unlike earlier-type bulge-dominated galaxies, which form at earlier times and therefore conform to the standard expectations of ΛCDM cosmology. This is because the central mass densities of galaxies tend to reflect the density of the Universe at their formation time (Wechsler et al., 2002). Nevertheless, while this explanation may be appropriate for the lower central densities typically found in later-type galaxies, it does not explain the shape of the inner density profiles, as the N-body simulations presented by Wechsler et al., (2002) still result in central cuspy densities for all mass scales.

The last possible explanation of this discrepancy between theory and observations is that the simulations are based on pure dark matter, and do not contain any baryonic physics. If one were to include baryonic physics in the simulations, the baryons may interact with the dark matter in some way to resolve the cusp/core problem. However, the effect of

baryons on the dark matter halo may work to make the problem even worse. As early as the mid-1980s it was shown that as baryons cool and fall into the center of a dark matter halo to form a visible galaxy, the halo contracts adiabatically. This makes the central density profile of the halo even steeper and sets up other problems, such as the Tully-Fisher zeropoint problem which we discuss further in section 4. Nevertheless, feedback from star formation has been suggested as a method to transform the halo profile from a cuspy, NFW profile, to a more shallow power law or constant density core. If feedback from star formation can affect the early stages of halo formation, steep cuspy density profiles may be transformed into the observed flat cores (Governato et al., 2010). Governato et al., (2010) use hydrodynamical simulations (in the framework of ΛCDM cosmology) in which the inhomogeneous interstellar medium is resolved. In their simulations, strong outflows from supernovae remove low-angular-momentum gas, which inhibits the formation of bulges and decreases the dark-matter density to less than half of what it would otherwise be within the central kiloparsec. Galaxies that are bulgeless and have shallow central dark-matter profiles arise from their simulations. These are the analogues of the low surface brightness and dwarf galaxies that are observed to have constant density core profiles in their central regions. Fig. 5 shows a result of the Governato et al., (2010) simulations. A comparison is made between a simulated dwarf galaxy rotation curve and an observed rotation curve of a real dwarf galaxy. The simulations by Governato et al., (2010) seem to provide a working solution to the cusp/core problem. From their work, it appears that supernova feedback provides a mechanism for transforming NFW density profiles into shallower power-law central density profiles. However, even in the case of the simulations by Governato et al., (2010), all of the observational studies that have dealt with the cusp/core problem have focused on the properties of late-type dwarf and low-surface brightness galaxies (Gentile et al., 2004, 2005; Kuzio de Naray et al., 2006, 2008; Shankar et al., 2006; Simon et al., 2005; Spano et al., 2008; Swaters et al., 2003). Galaxies with dominant bulges have yet to be explored in detail. In the few cases where such galaxies have been explored, a cuspy profile seems to result from the observational data (Seigar et al., 2008; Tamm et al., 2007; Tempel et al., 2007). Also, in the case of elliptical galaxies (which could be thought of as pure-bulge galaxies) and clusters of galaxies the NFW model provides a remarkable fit to the density profile down to ∼1% of the dark matter halo's virial radius (Humphrey et al., 2006; Zappacosta et al., 2006). This highlights the need to extend these studies to include galaxies with significant bulges. In the past, such studies have been difficult because the central regions of these galaxies are dominated by baryonic matter (unlike late-type systems which are dark matter dominated over large radial ranges). Determining the central baryonic mass depends on several factors such as signal-to-noise ratio and the accuracy of stellar mass-to-light ratios derived from stellar population synthesis codes (e.g., Maraston 2005). With new imaging surveys of nearby galaxies providing exquisite data quality and extremely good signal-to-noise ratio in the central regions of galaxies, it is likely that in the next decade these studies will be extended to bulge-dominated galaxies. One such survey is the Carnegie-Irvine Galaxy Survey (CGS; Ho et al., 2011) which consists of deep $BVRIK_s$ imaging of the 605 brightest southern hemisphere galaxies and will also include integral field spectroscopy of every galaxy in the southern hemisphere with a B band magnitude $B_T \leq 12.9$.

To complement these observational studies, simulations of the effects of supernova feedback (similar to those of Governato) on the halos of bulge-dominated galaxies are necessary. Since there is more stellar mass in the central regions of these systems, one might naively expect there to be a higher star formation rate and therefore a higher supernova rate in the histories of such galaxies. Presumably, this would have the effect of making central cores larger, yet
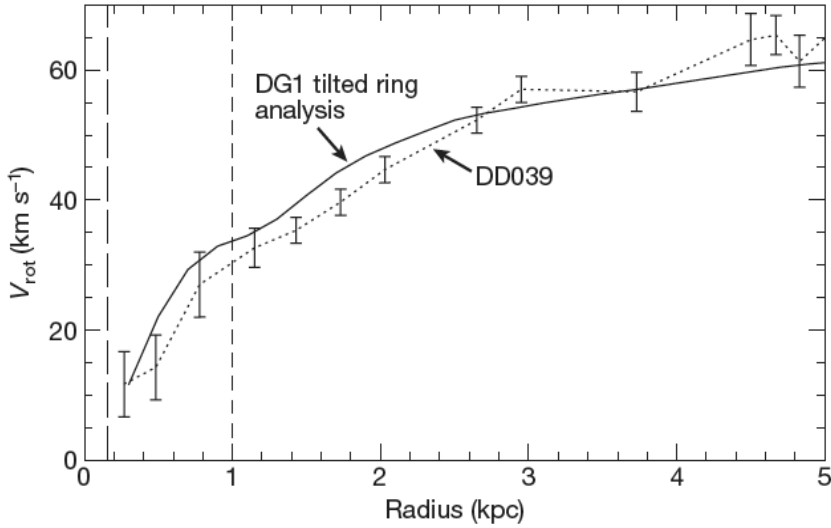
Fig. 5. Simulated rotation curve of a dwarf galaxy from Governato et al., (2010) compared to that measured for a real galaxy. The continuous black line shows the rotation velocity of the galaxy using the actual projected velocity field and the tilted ring analysis (Valenzuela et al., 2007). The dotted line shows the rotation curve of the galaxy DDO 39 as measured using a similar technique (Swaters et al., 2003), with standard deviation error bars. The velocity profile of both the observed and simulated galaxies imply a dark-matter distribution with a core scale length of about one kiloparsec, as directly measured in the simulation. The long-dashed vertical line shows the force resolution of the simulation, whereas the short-dashed vertical line marks the approximate scale length of the dark-matter "core". The underlying dark matter density is proportional to $r^{-\alpha}$, with $\alpha = 0.6$ in the central kiloparsec, consistent with observational estimates and shallower than a dark-matter-only simulation that would predict a steeper profile with $\alpha = 1.3$.

the observational data that does exist (admittedly for just a handful of bulge-dominated objects) suggests that these systems have central density cusps (e.g., Klypin et al., 2002; Seigar et al., 2008). It is therefore important to actually undertake a program of simulations to see how supernova feedback effects the central density profiles of dark matter halos of bulge-dominated galaxies.

4. The Tully-Fisher zero-point problem

As visible matter or baryons cool and fall into the center of a dark matter halo, an adiabatic invariant is set up and this leads to the adiabatic contraction of the halo. This so-called "adiabatic contraction" was used as early as Eggen et al., (1962) to explain the formation of the Milky Way galaxy, albeit without a dark matter halo. In the modern context of ΛCDM cosmology, the idea is that as baryons cool, they sink dissipatively to the center of a halo, decreasing there radial extent by about a factor of 10. Using this idea of baryonic infall, Blumenthal et al., (1986) showed that the underlying dark matter distribution is strongly perturbed. The dark matter distribution gets pulled inward, creating denser central profiles than would have evolved without dissipation. Blumenthal et al., (1986) found that under such

dissipational collapse, baryonic infall yields rotation curves that are flat over large distances, consistent with observations of spiral galaxies (e.g., Rubin et al., 1980, 1985). Blumenthal et al. (1986) assume that the initial spherically symmetric mass distribution, $M_i(r)$, represents a dynamical equilibrium state with a constant fraction, F, of dissipational baryons as a function of r. The dissipational particles then cool and fall into a final mass distribution, $M_b(r)$, which, in the case of a spiral galaxy, is constrained by the initial angular momentum distribution. The adiabatic invariant of the dissipationless particle orbits implies

$$r[M_b(r) + M_x(r)] = r_i M_i(r_i) = \frac{r_i M_x(r)}{(1 - F)}, \tag{2}$$

which can be solved iteratively for the final dark-matter mass distribution $M_x(r)$ given the initial total mass distribution $M_i(r_i)$ and the final mass distribution $M_b(r)$. The parameter r_i is the initial orbital radius, and $M_x(r) = (1 - F)M_i(r_i)$, because dissipationless particle orbits cannot cross. Blumenthal et al., (1986) assume that the final baryonic mass distribution is of the form $M_b(r) \propto 1 - (1 + r/b)\exp(-r/b)$, which describes the radial mass distribution of a thin disk whose density in the plane of the disk decreases exponentially with scale length b.

The resulting model rotation curves from Blumenthal et al., (1986) are shown in Fig. 6, which shows curves for various values of F and b, assuming there is no baryonic infall and therefore no change in the dissipationless matter beyond $r = 1$. They showed that for $F = 0.1$ and $b = 0.07$, the rotation curve is flat beyond $\sim$2 disk scale lengths, even though the initial rotation curve is not flat. One can also see that for $F > 0.1$, the rotation curve declines after $\sim$2 disk scale lengths, while for $F < 0.1$, dissipation produces too little change in the rotation curve. Blumenthal et al., (1986) therefore conclude that their choice of b requires that $0.05 < F < 0.2$ to produce observed rotation curves.

In ΛCDM simulations of structure formation, as galaxies form, and adiabatic invariant (similar to that shown in equation 2) is setup as a natural consequence of baryon infall, and this leads to the contraction of the host dark matter halo (e.g., Bullock et al., 2001; Gnedin et al., 2004). However, this leads to a problem with ΛCDM cosmology. Such models cannot reproduce the relation between galaxy luminosity and circular velocity (Tully & Fisher, 1977; Fig. 7) without over-producing the number density of galaxies at fixed luminosity (Benson et al., 2003; Cole et al., 2000; Gonzalez et al., 2000; Yang et al., 2003). This is known as the Tully-Fisher zero-point problem.

The Tully-Fisher relation can be used as a method to determine distances to galaxies. If we can determine the rotational velocity of a galaxy, we know the absolute magnitude of the galaxy from the Tully-Fisher relation, so given its apparent magnitude, the distance to the galaxy can be determined. Before the discovery of this relation, astronomers typically used the virial theorem to determine galaxy distances. This method was first adopted by Opik (1922), who used the virial theorem to determine the distance to M31. The argument is simple. A self-gravitating system will obey the law

$$M = kW^2 R/G \tag{3}$$

where M is the mass, W is a characterization of the motions in the body with units of speed and dominated by rotation in the case of a disk galaxy, R is a measure of the linear size, G is the universal gravitational constant, and k is a dimensionless constant of order unity that depends on the geometry of the system. Opik (1922) took a dependence of the mass $M \propto (M/L)ld^2$, where M/L is the mass-to-light ratio, l is the observed brightness and d is the observed distance. The virial theorem continued to be used to determine galaxy distances into
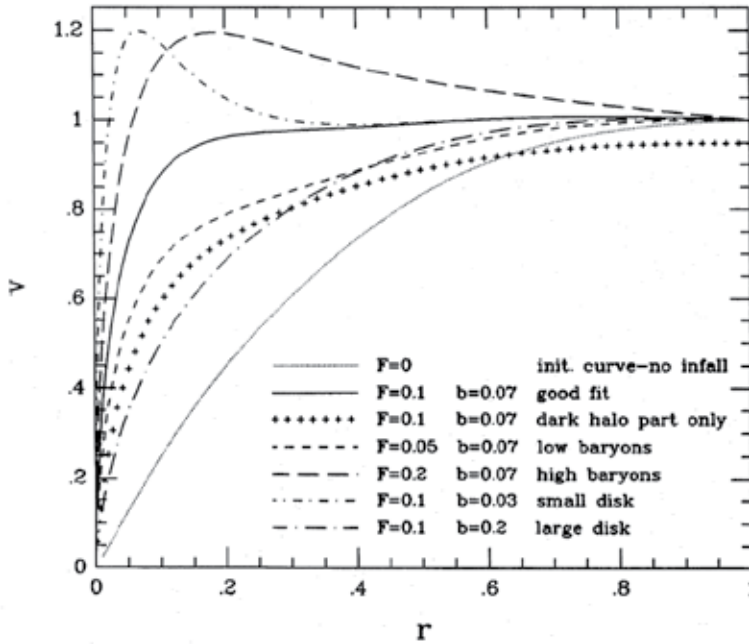
Fig. 6. A plot of the rotational velocity, $v = (M/R)^{1/2}$, versus radius, r, for the adiabatic invariant given in equation 2. The baryon fraction F is assumed to dissipate and fall into an exponential disk with scale length b. No dissipation is assumed to occur beyond $r = 1$. All curves except the plusses correspond to a total rotational velocity; the curve denoted by plus signs is the rotational velocity due to the dissipationless particles (i.e., the dark matter) alone for the case where $F = 0.1$ and $b = 0.07$.

the 1970s, with what was usually referred to as the "indicative mass" method. It was stripped down to the form $L \propto W^2 R$. With the discovery of the Tully-Fisher relation, Tully & Fisher (1977) suggested the use of two alternatives:

$$L \propto W^{\alpha} \tag{4}$$

$$L \propto R^{\beta}. \tag{5}$$

These alternatives reduce the number of variables and allow departures from the virial theorem. Equation 4 has the attraction that only one parameter depends on distance. It is this equation that became known as the Tully-Fisher Relation.

The motivation for the development of the relationship was based on well founded physics: more massive galaxies would be both more luminous and rotate faster. It was appreciated that there was a lot of leverage for the measurement of distances because of the d^2 dependence of observed luminosity. However the small scatter and apparent universality of the relationship could not have been fully anticipated. These properties have made the relationship interesting not only for the determination of distances but also as a constraint on models of the formation and evolution of galaxies. In particular, note that Opik's derivation of the virial relation above assumes a universal value for the mass-to-light ratio, M/L. Given the dominance of dark matter in spiral galaxies, it is not at all obvious that there would be a universal relationship between mass and light. The fact that the Tully-Fisher relation has as little scatter as it does is

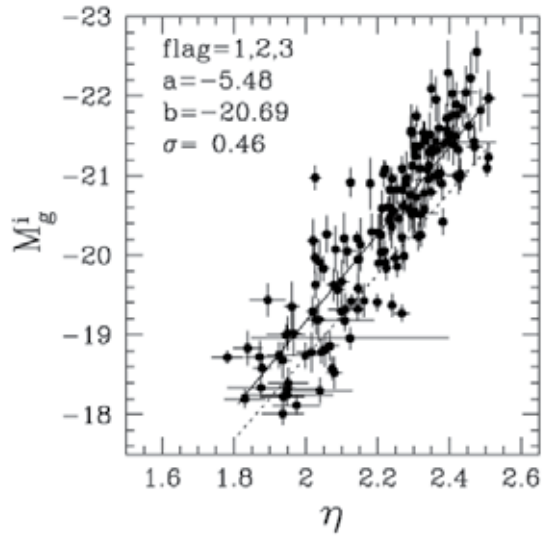Fig. 7. Tully-Fisher relation from Pizagno et al., (2007). The plot shows the internal-extinction corrected g-band magnitude (M_g^i) as a function of the inclination-corrected maximum rotation velocity ($\eta = \log(v)/\sin i$, where i is the inclination angle to the line of sight). The solid line is the best fit Tully-Fisher relation. The dotted line shows the best-fit Tully Fisher relation without correction for internal extinction.

seen as an important clue regarding the connection between the luminous and non-luminous components of spirals.

The fact that standard cosmological models cannot reproduce the Tully Fisher relation without over-producing the number density of galaxies at fixed luminosity is a fundamental problem for our cosmological models. Recently, Dutton et al., (2007) and Gnedin et al., (2007) revisited this problem using two large, well-defined, samples of disk-dominated (late-type) galaxies. They both took, as a starting point, the "standard" model of disk galaxy formation, which assumes that NFW dark halos respond to the formation of a disk via adiabatic contraction or AC (Blumenthal et al., 1986; Choi et al., 2006; Gnedin et al., 2004; Sellwood & McGaugh, 2005). Dutton et al., (2007) used the observed Tully-Fisher relation and the observed size-luminosity relation to single out a specific scenario for disk galaxy formation in ΛCDM cosmology. Their models involved four independent lognormal random variables: dark halo concentration $c_{\rm vir}$, disk spin $\lambda_{\rm gal}$, disk mass fraction $m_{\rm gal}$, and stellar mass-to-light ratio Γ_I. In order to get a simultaneous match of the Tully-Fisher and size-luminosity zero points with adiabatic contraction, Dutton's models require low-concentration halos, but this model has $V_{2.2} \sim 1.8V_{\rm vir}$ (where $V_{2.2}$ and $V_{\rm vir}$ are the circular velocity at 2.2 disk scale lengths and the virial radius, respectively), which is unable to match the luminosity function of galaxies (the observed ratio, $V_{2.2}/V_{\rm vir}$ as a function of halo concentration, as demonstrated in Dutton et al., (2007) is shown in Fig. 8). Similarly Dutton's models without adiabatic contraction but standard $c_{\rm vir}$ also predict high values of $V_{2.2}/V_{\rm vir}$ (see Fig. 9). Dutton et al., (2007) concluded that models in which disk formation induces an expansion rather than the commonly assumed contraction of the dark matter halos have $V_{2.2} \sim 1.2V_{\rm vir}$, which allows a simultaneous fit of the luminosity function (see Fig. 10). This may result from non-spherical, clumpy gas accretion, where dynamical friction transfers energy from the gas to the dark matter. This model requires

low λ_{gal} and m_{gal} values, contrary to naive expectations. However, the low λ_{gal} is consistent with the notion that disk galaxies predominantly survive in halos with a quiet merger history, while a low m_{gal} is also indicated by galaxy-galaxy lensing. The smaller than expected scatter in the radius-luminosity relation and the lack of correlation between the residuals of the Tully-Fisher and radius-luminosity relations (found in Dutton's expansion models), respectively, imply that the scatter in λ_{gal} and in c_{vir} needs to be smaller than predicted for ΛCDM halos. Again this is consistent with the idea that disk galaxies preferentially reside in halos with a quiet merger history.
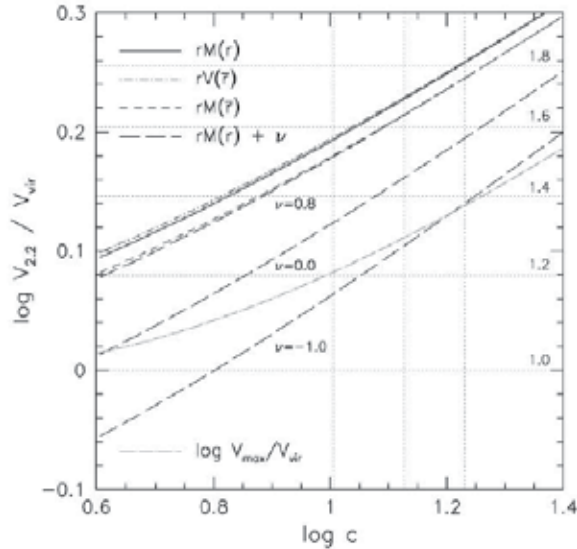
Fig. 8. Ratio of observed to virial circular velocities, $V_{2.2}/V_{\text{vir}}$, as a function of the concentration parameter, for various forms of adiabatic contraction. All models have $\lambda_{\text{gal}} = 0.048$ and $m_{\text{gal}} = 0.05$. The vertical dotted lines show the mean concentration parameter according to the model of Bullock et al., (2001) for halos with virial mass, $M_{\text{vir}} = 10^{13}, 10^{12}$, and $10^{11}\ h^{-1}\ M_{\odot}$. The horizontal dotted lines show $V_{2.2}/V_{\text{vir}}$ ratios of 1–1.8 at intervals of 0.2. The relation using the standard Blumenthal et al., (1986) adiabatic invariant, $rM(r)$, is given by the solid line. The modified adiabatic invariant, $rM(\bar{r})$, as proposed by Gnedin et al., (2004), results in only a $\simeq 0.02$ dex ($\simeq 5\%$) reduction in $V_{2.2}$ (short-dashed line). Most of this reduction is taken back if specific angular momentum is used as the adiabatic invariant and the disk geometry is taken into account when computing V_{circ} (dot-short-dashed line). For all of these adiabatic contraction models, standard concentration parameters yield $V_{2.2}/V_{\text{vir}} \simeq 1.6$. The long-dashed lines show models where the contraction factor has been artificially modified by Dutton et al., (2007). Starting from the standard model ($\nu = 1$), the $rM(\bar{r})$ relation is approximately reproduced with $\nu = 0.8$, $\nu = 0$ results in no adiabatic contraction of the halo, while $\nu = 1$ gives halo expansion. The dot-long-dashed line shows $V_{\text{max}}/V_{\text{vir}}$, where V_{max} is the maximum halo circular velocity (without adiabatic contraction).

In contrast, Gnedin et al., (2007) investigate the structural properties of dark matter halos of disk galaxies, using a well-defined sample of 81 disk-dominated galaxies from the Sloan Digital Sky Survey (SDSS). They model the mass Tully-Fisher and fundamental plane relations

Fig. 9. Tully-Fisher and size-luminosity relations using fiducial mean c_{vir}, λ_{gal}, and m_{gal} parameters and different values of mass-to-light ratio, Γ_I, Toomre stability parameter, Q, and galaxy mass fraction slope, $\al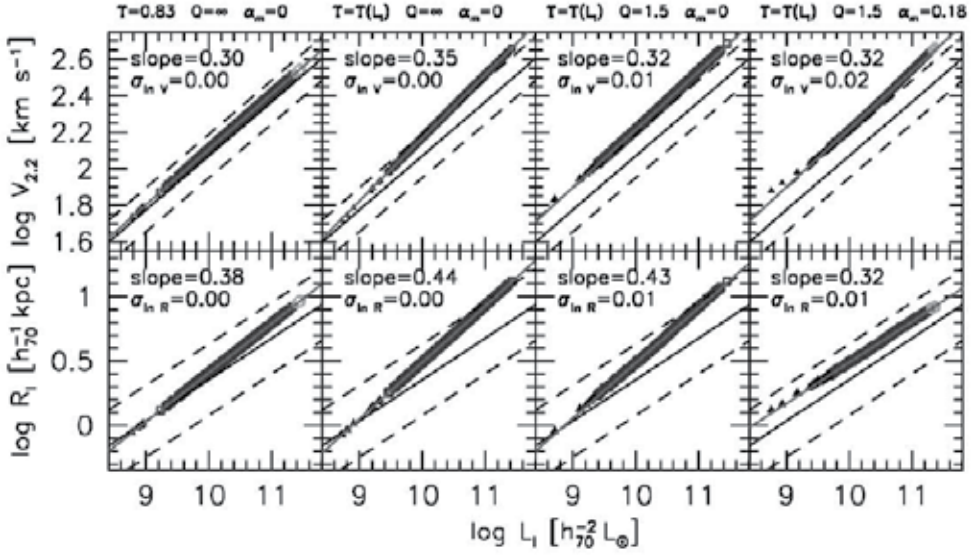pha_m$, from Dutton et al., (2007). The black solid and dashed lines show the mean and 2-σ scatter of the observations. The grey line is a fit to the model galaxies, whose slope is given in the top left corner of each box. The model on the far left is the Mo et al., (1998) model with $\Gamma_I = 0.83$, $Q = \infty$ (a pure stellar disk), and $\alpha_m = 0$. The model in the left middle only differs from the left model in the value of Γ_I. This steepens the slopes of both relationships, making them incompatible with the data. The right middle model includes the star formation threshold, which results in a gas fraction that increases with decreasing luminosity and hence a shallower slope for the Tully-Fisher relationship, in agreement with the data. The effect on the slope of the size-luminosity relationship is not as strong because as the luminosities decrease, the stellar disk scale-lengths also decrease. In order to match the size-luminosity relationship slope, Dutton et al., (2007) set $\alpha_m = 0.18$ (far right panel). While this model matches the slope of the Tully-Fisher and size-luminosity relationships, it fails to reproduce the zero-points.

of these galaxies, using the observed stellar mass, M_*, the disk scale-length, R_d, and the optical Hα rotation velocity at $2.2R_d$ ($V_{2.2}$). Gnedin et al., (2007) include the effects of adiabatic contraction in their models, arguing that it is a fundamental prediction in galaxy formation. They find that models with a constant stellar disk-to-total mass fraction, m_d, under-predict the intrinsic scatter in the Tully-Fisher and fundamental plane relations. Introducing a scatter in m_d and allowing the mean value, $\overline{m_d}$, to significantly scale with the stellar surface density, improves the observational match. The distribution of angular momentum parameters required to match the observed scale-lengths is an output of the Gnedin models and is narrower than that predicted for halo spin. However, the best-fit models from Gnedin et al., (2007), with a Kroupa (2001) initial mass function (IMF), overproduce the galaxy stellar mass function and predict virial mass-to-light ratios lower than those inferred from galaxy-galaxy weak lensing (Hoekstra et al., 2005; Mandelbaum et al., 2006) and satellite dynamics (Prada et al., 2006). Gnedin et al., (2007) therefore argue that the IMF needs to be modified by adopting a "light" stellar IMF with M_*/L lower by 0.15 dex. Adopting such an IMF reproduces the
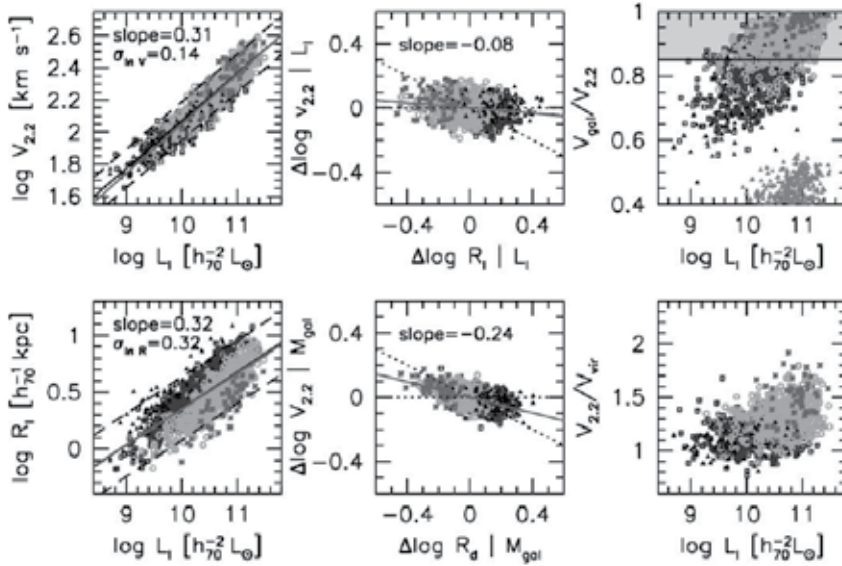
Fig. 10. Scaling relations and residual correlations of Dutton's halo expansion model from Dutton et al., (2007). This model reproduces the slopes, scatter, and zero points of the Tully-Fisher and size-luminosity relations. This model has a median $V_{2.2}/V_{\rm vir} = 1.22$, which is often required in order to simultaneously fit the Tully-Fisher and size-luminosity relations.

slope and zero point of the Tully Fisher relation and other fundamental plane relations (see Fig. 11).
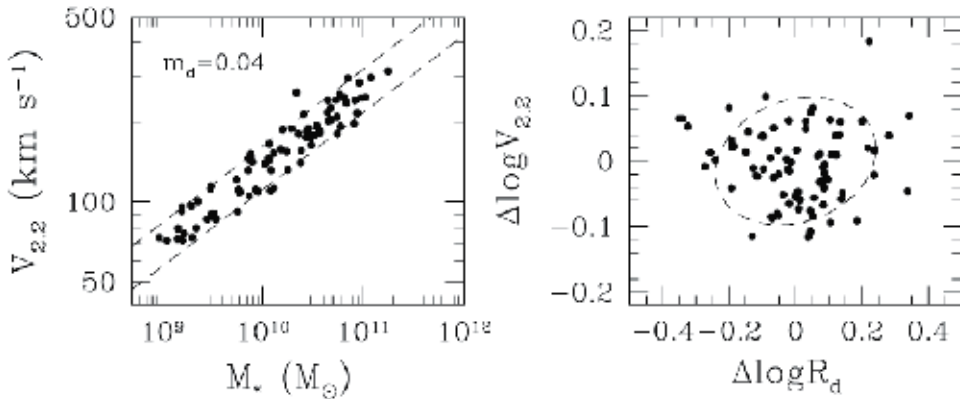
Fig. 11. Best fit models from Gnedin et al., (2007) to the Tully-Fisher relation for the case of a "light" IMF, including adiabatic contraction. The dashed lines representing the observations in the left panel and shifted by -0.15 dex in M_* to account for the modified IMF.

More recently, Dutton et al., (2010) have investigated the origin of the relations between stellar mass and optical circular velocity for early-type and late-type galaxies — the Faber-Jackson and Tully-Fisher relations. They combine measurements of dark halo masses (from satellite kinematics and weak lensing) and the distribution of baryons in galaxies (from a new compilation of galaxy scaling relations) with constraints on dark halo structure from

cosmological simulations. The principle unknowns are the halo response to galaxy formation and the stellar IMF. The slopes of the Tully-Fisher and Faber-Jackson relations are naturally reproduced for a wide range of halo response and IMFs within Dutton's models. However, models with a universal IMF and universal halo response cannot simultaneously reproduce the zero points of both relations. For a model with a universal Chabrier (2003) IMF, late-type galaxies require halo expansion, while early-type galaxies require halo contraction (see Fig. 12). A Salpeter (1955) IMF is permitted for high mass ($\sigma > 180$ km/s) early-type galaxies, but is inconsistent for intermediate masses, unless $V_{circ}(R_e)/\sigma_e > 1.6$. If the IMF is universal and close to the Chabrier IMF, Dutton et al., (2010) speculate that the presence of a major merger may be responsible for the contraction in early-type galaxies while clumpy accreting streams and/or feedback leads to expansion in late-type galaxies. Alternatively, a recently proposed variation in the IMF disfavors halo contraction in both types of galaxies. The models from Dutton et al., (2010) naturally reproduce flat and featureless circular velocity profiles within the optical regions of galaxies without fine-tuning.
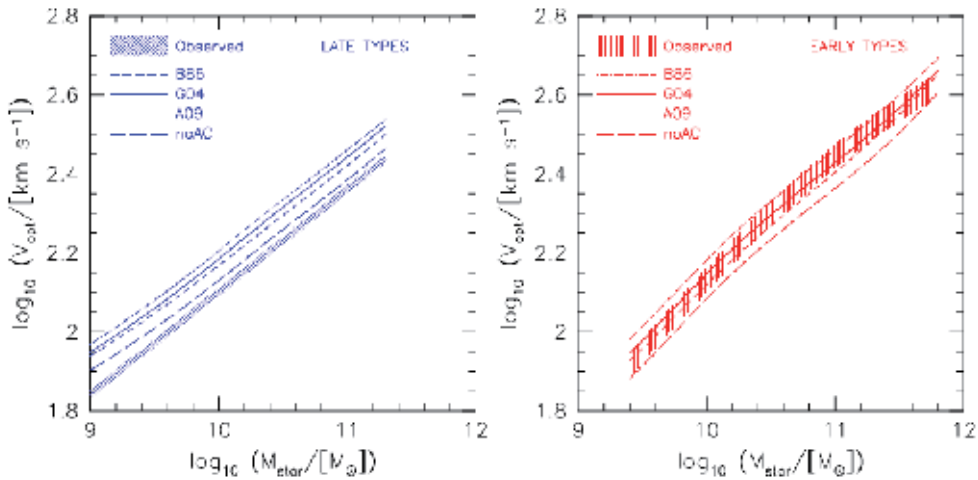
Fig. 12. Effect of adiabatic contraction on the circular velocity (V_{opt}) versus stellar mass (M_{star} relation for early type (right panel) and late-type (left panel) galaxies using the models of Dutton et al., (2010). Adiabatic contraction results in higher optical circular velocities. Compared to a model with no halo contraction (long dashed lines), the increase in optical circular velocities is $\simeq 0.1$ dex for the Blumenthal et al., (1986) model, $\simeq 0.08$ dex for the Gnedin et al. (2004) model, and $\simeq 0.05$ dex for the Abadi et al., (2010) halo contraction model. For these models (with a Chabrier IMF and standard halo concentrations from Maccio et al., 2008) early-type galaxies favor models with strong halo contraction, whereas late-type galaxies favor models with halo expansion.

The idea that early-type bulge-dominated disk galaxies (Dutton et al., 2010) require halo contraction is consistent with observational evidence that suggests that the halo of our nearest large neighbor (M31, a bulge-dominated galaxy) has a rotation curve that is consistent with a dark matter halo that has undergone adiabatic contraction (Klypin et al., 2002; Seigar et al., 2008). Seigar et al., (2008) model the dark matter contribution to the rotation curve of M31 using several different models for adiabatic contraction and a model that includes no halo contraction. They find that the halo model that best reproduces the rotation curve of M31,

with a halo concentration in the expected range (Maccio et al., 2008) requires a halo model with the strongest version of adiabatic contraction (see Fig. 13).
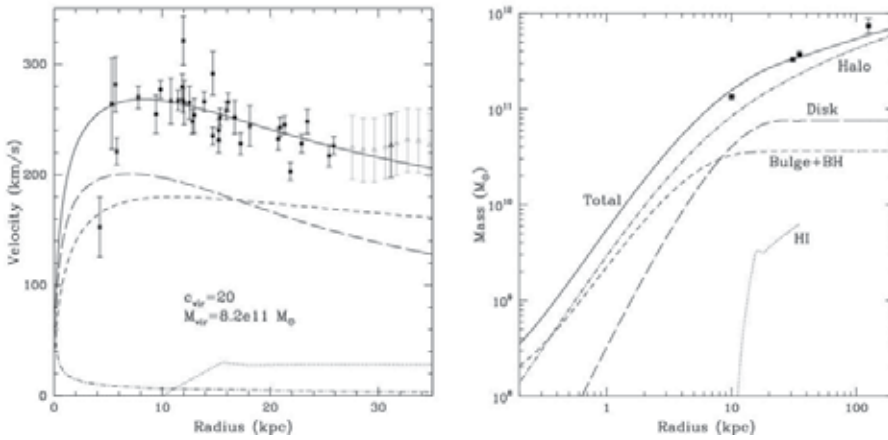
Fig. 13. Left panel: Rotation curve data (solid square points) from Rubin & Ford, (1970) with the best-fitting model, with adiabatic contraction overlaid (solid line). For comparison the HI rotation velocities (grey open triangles) from Carignan et al., (2006) are shown (the black point corresponds to the average velocity and radius of the Carignan data). The best-fitting rotation curve is decomposed into four components, the contribution from dark matter (short-dashed line), the contribution from baryonic matter (long-dashed line), the HI contribution (dotted line) from Carignan et al., (2006) and the contribution from the central $10^8 \, M_\odot$ black hole (dot-dashed line) from Bender et al., (2005). Right panel: Enclosed mass as a function of radius for the best-fitting rotation curve model with adiabatic contraction. The solid line indicates the total mass. Also indicated are the HI mass out to a 35 kpc radius (dotted line), the bulge+BH mass (short-dashed line), the disk mass (long-dashed line), and the halo mass (dot-dashed line). The data points indicate masses derived using other observational methods at fixed radii of 10 kpc, 31 kpc, 35 kpc, and 125 kpc from Rubin & Ford, (1970), Evans & Wilkinson, (2000), Carignan et al., (2006), and Fardal et al., (2006) respectively. Both figures are taken from Seigar et al., (2008).

On the contrary, disk-dominated galaxies, with little or no bulge, should undergo a halo expansion according to the models of Dutton et al., (2007, 2010). However, observations of disk-dominated galaxies, including M33, suggest that a pure NFW model is sufficient to describe the dark matter halo contribution to their rotation curves (Kassin et al., 2006; Seigar 2011). For a sample of 34 bright spiral galaxies, Kassin et al., (2006) compared their rotation curve data to the NFW parametrization for dark matter profiles with and without including adiabatic contraction. All but two of their galaxies are better fit if adiabatic contraction does not occur. However, their fits are generally poor, and the possibility that the fits may improve if a halo expansion model is applied cannot be ruled out.

Finally, Klypin et al. (2010) claim to have run simulations that reproduce the Tully-Fisher relation slope and zero-point without the need for halo contraction or expansion. Their method consists of matching the abundance of galaxies in the local Universe to that of halos in their simulation. Their simulation (termed the Bolshoi simulation) was done in a volume 250 h^{-1} Mpc on a side using 8 billion particles with mass and force resolution adequate to follow subhalos down to a completeness limit of $V_{\rm circ} = 50 \; {\rm km \, s^{-1}}$ maximum circular velocity. The

halos in this simulation are matched to the abundance of galaxies in the local Universe by using the triple-Schechter fit to the galaxy luminosity function obtained from the Sloan Digital Sky Survey Data Release 6 galaxy sample by Montero-Dorta & Prada, (2009) and assigning r-band absolute magnitudes to their dark matter halos starting with the largest until they reach the maximum circular velocity completeness limit of the simulation. This assumption is reasonable for halos of Milky Way-sized galaxies or smaller, but it is not adequate for larger galaxies. For larger, brighter galaxies, the circular velocities are overestimated. Fig. 14 shows the Tully-Fisher relation plot for all halos in the Bolshoi simulation, compared with several observational data sets. The plotted V_{circ} is measured at 10 kpc and includes the effect of baryons estimated to have reached the central regions of galaxies, but does not include the effects of halo contraction. The agreement with the galaxy data is remarkably good from $V_{circ} \sim 50$ km s^{-1} to $V_{circ} \sim 200$ km s^{-1}. At higher V_{circ} Klypin et al., (2010) claim that an increasing fraction of halos host early-type galaxies, groups, or clusters (although it should be noted that M31 has a $V_{circ} \simeq 270$ km s^{-1} and observational data suggest halo contraction is required for M31). Given the lack of disk galaxies with $V_{circ} > 300$ km s^{-1}, a bright elliptical galaxy, M87, is included in Fig. 14. The circular velocity was obtained by the mass modeling of Strom et al., (1981) and the r-band magnitude was found by integrating the surface brightness profile of Strom et al., (1981). The result of the method presented by Klypin et al., (2010) is that the Tully-Fisher relation is reproduced without the need for contraction or expansion. Nevertheless, Fig. 14 seems to show that the datapoints all seem to lie systematically above the models produced by Klypin et al., (2010). It also needs to be demonstrated that this model can reproduce the rotation curves of individual galaxies.
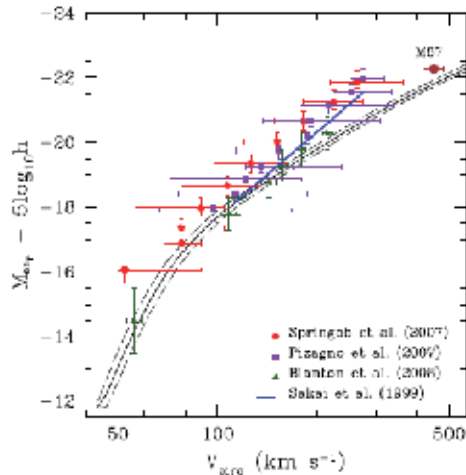
Fig. 14. Comparison of the Tully-Fisher relation of different observational datasets with the theoretical predictions of "abundance-matching" produced in Klypin et al., (2010). Median and 90% contours of r-band magnitudes versus circular velocity of the mock galaxies in the Bolshoi simulation are shown.

Given the fundamental issues at hand, the question of whether adiabatic contraction or a halo expansion occurs in nature is of significant interest. Most simulations favor the idea that halos contract, especially when the infall of baryons is smooth and adiabatic (Blumenthal et al., 1986; Ryden & Gunn, 1987; Sellwood & McGaugh, 2005). Simulations also suggest that halos will contract even when galaxies (or clusters) form quickly from an irregular collapse (Barnes,

1987; Choi et al., 2006; Flores et al., 1993; Gnedin et al., 2004; Jesseit et al., 2002; Sellwood & McGaugh, 2005; Weinberg et al., 2008). However, there is the possibility that halos (especially of disk-dominated galaxies) expand (Dutton et al., 2010). Furthermore, a new approach where the galaxy abundances are matched based on simulated halos and subhalos seems to alleviate the need for any change (either expansion or contraction) of the halo density profile (Klypin et al., 2010). Few of these relatively new ideas have been tested. To date, observational studies that have dealt with the Tully-Fisher zero-point problem have all focused on the properties of late-type spiral galaxies (Hubble types Sc and later, i.e., galaxies with little or no bulge) or dwarf galaxies (Kassin et al., 2006; Pizagno et al., 2005, 2007). When galaxies with dominant bulges have been explored, a contracted NFW halo does seem to fit the data, as is the case for M31 (see Fig. 13; Klypin et al., 2002; Seigar et al., 2008; Tamm et al., 2007; Tempel et al., 2007). This highlights the need to extend these studies to include galaxies with significant bulges. Once again, for such a study, surveys such as the CGS (Ho et al., 2011) will prove invaluable.

5. Conclusions

In conclusion, ΛCDM cosmology reproduces the large scale structure of the Universe extremely accurately (Blumenthal et al., 1984; Springel et al., 2005). However, on galaxy-sized scales there are several known issues with the theory. Here, we have highlighted two particular problems. The first, namely the cusp/core problem, highlights the fact that ΛCDM simulations of structure formation predict central cusps in the density profiles of dark matter halos. However, observations, particularly of LSB galaxies, suggest that the central densities are flat over the inner $\sim$1 kpc. This is a serious issue with ΛCDM cosmology. A solution involving supernova feedback may be the answer (e.g., Governato et al., 2010), but this model needs to be applied to large spiral and elliptical galaxies to see how widely applicable it is.

The second problem, the Tully-Fisher zeropoint problem, highlights the fact that ΛCDM simulations cannot reproduce the Tully-Fisher relation without over-producing the number density of galaxies at fixed luminosity. Recently, Klypin et al., (2010) claim to have solved this issue, but it seems to systematically overestimate galaxy rotation velocities, and so further study of their model is required.

6. Acknowledgements

7. References

Bailin, J., Simon, J. D., Bolatto, A. D., Gibson, B. K. & Power, C. (2007) Self-Consistent Massive Disks in Triaxial Dark Matter Halos. *The Astrophysical Journal*, Vol. 667, 191–201

Barnes, J. E. (1987) Halo Response to Galaxy Formation. In: *Nearly Normal Galaxies: From the Planck Time to the Present*, Faber, S. M., (Ed.), 154–159, Springer-Verlag, 978-0-387-96521-1, New York

Bender, R., Kormendy, J., Bower, G., Green, R., Thomas, J., Danks, A. C., Gull, T., Hutchings, J. B., Joseph, C. L., Kaiser, M. E., Lauer, T. R., Nelson, C. H., Richstone, D., Weistrop, D. & Woodgate, B. (2005) HST-STIS Spectroscopy of the Triple Nucleus of M31: Two Nested Disks in Keplerian Rotation around a Supermassive Black Hole. *The Astrophysical Journal*, Vol. 631, 280–300

Benson, A. J., Bower, R. G., Frenk, C. S., Lacey, C. G., Baugh, C. M. & Cole, S. (2003) What Shapes the Luminosity Function of Galaxies? *The Astrophysical Journal*, Vol. 599, 38–49

Bertschinger, E. (1985) Self-Similar Secondary Infall and Accretion in an Einstein-de Sitter Universe. *The Astrophysical Journal Supplement*, Vol. 58, 39–65

Blumenthal, G. R., Faber, S. M., Primack, J. R. & Rees, M. J. (1984) Formation of Galaxies and Large-Scale Structure with Cold Dark Matter. *Nature*, Vol. 311, 517–525

Blumenthal, G. R., Faber, S. M., Flores, R. & Primack, J. R. (1986) Contraction of Dark Matter Galactic Halos Due to Baryonic Infall. *The Astrophysical Journal*, Vol. 301, 27–34

Bullock, J. S., Kolatt, T. S., Sigad, Y., Somerville, R. S., Kravtsov, A. V., Klypin, A. A., Primack, J. R. & Dekel, A. (2001) Profiles of Dark Halos: Evolution, Scatter and Environment. *Monthly Notices of the Royal Astronomical Society*, Vol. 321, 559–575

Carignan, C., Chemin, L., Huchtmeier, W. K. & Lockman, F. J. (2006) The Extended H I Rotation Curve and Mass Distribution of M31. *The Astrophysical Journal Letters*, Vol. 641, L109–L112

Chabrier, G. (2003) Galactic Stellar and Substellar Initial Mass Function. *Publications of the Astronomical Society of the Pacific*, Vol. 115, 763–795

Choi, J.-H., Lu, Y., Mo H. J. & Weinberg, M. D. (2006) Dark Matter Halo Response to the Disc Growth. *Monthly Notices of the Royal Astronomical Society*, Vol. 372, 1869–1874

Cole, S., Lacey, C. G., Baugh, C. M. & Frenk, C. S. (2000) Hierarchical Galaxy Formation. *Monthly Notices of the Royal Astronomical Society*, Vol. 319, 168–204

Corbelli, E. & Salucci, P. (2000) The Extended Rotation Curve and the Dark Matter Halo of M33. *Monthly Notices of the Royal Astronomical Society*, Vol. 311, 441–447

Crone, M. M., Evrard, A. E. & Richstone, D. O. (1994) The Cosmological Dependence of Cluster Density Profiles. *The Astrophysical Journal*, Vol. 434, 402–416

de Blok, W. J. G., McGaugh, S. S. & van der Hulst, J. M. (1996) H I Observations of Low Surface Brightness Galaxies: Probing Low-Density Galaxies. *Monthly Notices of the Royal Astronomical Society*, Vol. 283, 18–54

de Blok, W. J. G., McGaugh, S. S. & Rubin, V. C. (2001) High-Resolution Rotation Curves of Low Surface Brightness Galaxies. II. Mass Models. *The Astronomical Journal*, Vol. 122, 2396–2427

de Vaucouleurs, G., de Vaucouleurs, A., Corwin, H. G., Buta, R. J., Paturel, G. & Fouqué, P. (1991). *The Third Reference Catalog of Bright Galaxies*, Springer-Verlag, 3-540-97552-7, New York.

Dubinski, J. & Carlberg, R. G. (1991) The Structure of Cold Dark Matter Halos. *The Astrophysical Journal*, Vol. 378, 496–503

Dutton, A. A., van den Bosch, F. C., Dekel, A. & Courteau, S. (2007) A Revised Model for the Formation of Disk Galaxies: Low Spin and Dark Halo Expansion. *The Astrophysical Journal*, Vol. 654, 27–52

Dutton, A. A., Conroy, C., van den Bosch, F. C., Prada, F. & More, S. (2010) The Kinematic Connection Between Galaxies and Dark Matter Halos. *Monthly Notices of the Royal Astronomical Society*, Vol. 407, 2–16

Efstathiou, G., Frenk, C. S., White, S. D. M. & Davis, M. (1988) Gravitational Clustering from Scale-Free Initial Conditions. *Monthly Notices of the Royal Astronomical Society*, Vol. 235, 715–748

Eggen, O. J., Lyndell-Bell, D. & Sandage, A. R. (1962) Evidence from the Motions of Old Stars that the Galaxy Collapsed. *The Astrophysical Journal*, Vol. 136, 748–767

Evans, N. W. & Wilkinson, M. I. (2000) The Mass of the Andromeda Galaxy. *Monthly Notices of the Royal Astronomical Society*, Vol. 316, 929–942

Fardal, M. A., Babul, A., Geehan, J. J. & Guhathakurta, P. (2006) Investigating the Andromeda Stream. II. Orbital Fits and Properties of the Progenitor. *Monthly Notices of the Royal Astronomical Society*, Vol. 366, 1012–1028

Fillmore, J. A. & Goldreich, P. (1984) Self-Similar Gravitational Collapse in an Expanding Universe. *The Astrophysical Journal*, Vol. 218, 1–8

Flores, R., Primack, J. R., Blumenthal, G. R & Faber, S. M. (1993) Rotation Curves from Baryonic Infall - Dependence on Disk-to-Halo Ratio, Initial Angular Momentum, and Core Radius, and Comparison with Data. *The Astrophysical Journal*, Vol. 412, 443–454

Flores, R. A. & Primack, J. R. (1994) Observational and Theoretical Constraints on Singular Dark Matter Halos. *The Astrophysical Journal*, Vol. 427, L1–L4

Frenk, C. S., White, S. D. M., Efstathiou, G. & Davis, M. (1985) Cold dark matter, the structure of galactic haloes and the origin of the Hubble sequence. *Nature*, Vol. 317, 595–597

Freank, C. S., White, S. D. M., Davis, M. & Efstathiou, G. (1988) The formation of dark halos in a Universe dominated by cold dark matter. *The Astrophysical Journal*, Vol. 327, 507–525

Gentile, G., Salucci, P., Klein, U., Vergani, D. & Kalberla, P. (2004) The cored distribution of dark matter in spiral galaxies. *Monthly Notices of the Royal Astronomical Society*, Vol. 351, 903–922

Gentile, G., Burkert, A., Salucci, P., Klein, U. & Walter, F. (2005) The Dwarf Galaxy DDO 47 as a Dark Matter Laboratory: Testing Cusps Hiding in Triaxial Halos. *The Astrophysical Journal*, Vol. 634, L34-L37

Gnedin, O. Y., Kravtsov, A. V., Klypin, A. A. & Nagai, D. (2004) Response of Dark Matter Halos to Condensation of Baryons: Cosmological Simulations and Improved Adiabatic Contraction Model. *The Astrophysical Journal*, Vol. 616, 16–26

Gnedin, O. Y., Weinberg, D. H.; Pizagno, J., Prada, F. & Rix, H.-W. (2007) Dark Matter Halos of Disk Galaxies: Constraints from the Tully-Fisher Relation. *The Astrophysical Journal*, Vol. 671, 1115–1134

Gonzalez, A. H., Williams, K. A., Bullock, J. S., Kolatt, T. S. & Primack, J. R. (2000) The Velocity Function of Galaxies. *The Astrophysical Journal*, Vol. 528, 145–155

Gottlber, S., Klypin, A. & Kravtsov, A. V. (2001) Merging History as a Function of Halo Environment. *The Astrophysical Journal*, Vol. 546, 223–233

Governato, F., Gardner, J. P., Stadel, J., Quinn, T. & Lake, G. (1999) On the Origin of Early-Type Galaxies and the Evolution of the Interaction Rate in the Field. *The Astronomical Journal*, Vol. 117, 1651–1656

Governato, F., Brook, C., Mayer, L., Brooks, A., Rhee, G., Wadsley, J., Jonsson, P., Willman, B., Stinson, G., Quinn, T. & Madau, P. (2010) Bulgeless dwarf galaxies and dark matter cores from supernova-driven outflows. *Nature*, Vol. 463, 203–206

Gunn, J. E. & Gott, J. R. (1972) On the Infall of Matter Into Clusters of Galaxies and Some Effects on Their Evolution. *The Astrophysical Journal*, Vol. 176, 1–19

Hayashi, E., Navarro, J. F. & Springel, V. (2007) The shape of the gravitational potential in cold dark matter halos. *Monthly Notices of the Royal Astronomical Society*, Vol. 377, 50–62

Hernquist, L. (1990) An analytical model for spherical galaxies and bulges. *The Astrophysical Journal*, Vol. 356, 359–364

Ho, L. C., Zhao-Yu, L., Barth, A. J., Seigar, M. S. & Peng, C. Y. (2011) Carnegie-Irvine Galaxy Survey. I. Overview and Atlas of Optical Images. *The Astrophysical Journal Supplement*, submitted

Hoekstra, H., Hsieh, B. C., Yee, H. K. C., Lin, H. & Gladders, M. D. (2005) Virial Masses and the Baryon Fraction in Galaxies. *The Astrophysical Journal*, Vol. 635, 73–85

Hoffman, Y. & Shaham, J. (1985) Local density maxima - Progenitors of structure. *The Astrophysical Journal*, Vol. 297, 16–22

Hoffman, Y. (1988) Angular momentum, hierarchical clustering, and local density maxima. *The Astrophysical Journal*, Vol. 329, 8–15

Humphrey, P. J., Buote, D. A., Gastaldello, F., Zappacosta, L., Bullock, J. S., Brighenti, F. & Mathews, W. G. (2006) A Chandra View of Dark Matter in Early-Type Galaxies. *The Astrophysical Journal*, Vol. 646, 899–918

Jesseit, R., Naab, T. & Burkert, A. (2002) The Validity of the Adiabatic Contraction Approximation for Dark Matter Halos. *The Astrophysical Journal*, Vol. 571, L89–L92

Kaplinghat, M., Knox, L. & Turner, M. S. (2000) Annihilating Cold Dark Matter. *Physical Review Letters*, Vol. 85, 3335–3338

Kassin, S. A., de Jong, R. S. & Weiner, B. J. (2006) Dark and Baryonic Matter in Bright Spiral Galaxies. II. Radial Distributions for 34 Galaxies. *The Astrophysical Journal*, Vol. 643, 804–824

Klypin, A., Zhao, H. & Somerville, R. S. (2002) ΛCDM-based Models for the Milky Way and M31. I. Dynamical Models. *The Astrophysical Journal*, Vol. 573, 597–613

Klypin, A., Trujillo-Gomez, S. & Primack, J. (2010) Halos and galaxies in the standard cosmological model: results from the Bolshoi simulation. *The Astrophysical Journal*, submitted (arXiv:1002.3660)

Komatsu, E., Smith, K. M., Dunkley, J., Bennett, C. L., Gold, B., Hinshaw, G., Jarosik, N., Larson, D., Nolta, M. R., Page, L., Spergel, D. N., Halpern, M., Hill, R. S., Kogut, A., Limon, M., Meyer, S. S., Odegard, N., Tucker, G. S., Weiland, J. L., Wollack, E. & Wright, E. L. (2011) Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. *The Astrophysical Journal Supplement*, Vol. 192, 18–64

Kroupa, P. (2001) On the variation of the initial mass function. *Monthly Notices of the Royal Astronomical Society*, Vol. 322, 231–246

Kuzio de Naray, R., McGaugh, S. S., de Blok, W. J. G. & Bosma, A. (2006) High-Resolution Optical Velocity Fields of 11 Low Surface Brightness Galaxies. *The Astrophysical Journal Supplement*, Vol. 165, 461–479

Kuzio de Naray, R., McGaugh, S. S. & de Blok, W. J. G. (2008) Mass Models for Low Surface Brightness Galaxies with High-Resolution Optical Velocity Fields. *The Astrophysical Journal*, Vol. 676, 920–943

Kuzio de Naray, R., Martinez, G. D., Bullock, J. S. & Kaplinghat, M. (2010) The Case Against Warm or Self-Interacting Dark Matter as Explanations for Cores in Low Surface Brightness Galaxies. *The Astrophysical Journal*, Vol. 710, L161–L164

Kuzio de Naray, R. & Kaufmann, T. (2011) Recovering cores and cusps in dark matter haloes using mock velocity field observations. *Monthly Notices of the Royal Astronomical Society*, in press (arXiv:1012.3471)

Lacey, C. & Cole, S. (1993) Merger rates in hierarchical models of galaxy formation. *Monthly Notices of the Royal Astronomical Society*, Vol. 262, 627–649

LeMaitre, G. (1931) Expansion of the Universe, A homogeneous Universe of constant mass and increasing radius accounting for the radial velocity of extra-galactic nebulae. *Monthly Notices of the Royal Astronomical Society*, Vol. 91, 483–490

Maccio, A. V., Dutton, A. A. & van den Bosch, F. C. (2008) Concentration, spin and shape of dark matter haloes as a function of the cosmological model: WMAP1, WMAP3 and WMAP5 results. *Monthly Notices of the Royal Astronomical Society*, Vol. 391, 1940–1954

Magrini, L., Corbelli, E. & Galli, D. (2007) The building up of the disk galaxy M33 and the evolution of the metallicity gradient. *Astronomy & Astrophysics*, Vol. 470, 843–855

Mandelbaum, R., Seljak, U., Cool, R. J., Blanton, M., Hirata, C. M. & Brinkmann, J. (2006) Density profiles of galaxy groups and clusters from SDSS galaxy-galaxy weak lensing. *Monthly Notices of the Royal Astronomical Society*, Vol. 372, 758–776

Maraston, C. (2005) Evolutionary population synthesis: models, analysis of the ingredients and application to high-z galaxies. *Monthly Notices of the Royal Astronomical Society*, Vol. 362, 799–825

Mo, H. J., Mao, S. & White, S. D. M. (1998) The formation of galactic discs. *Monthly Notices of the Royal Astronomical Society*, Vol. 295, 319–336

Montero-Dorta, A. D. & Prada, F. (2009) The SDSS DR6 luminosity functions of galaxies. *Monthly Notices of the Royal Astronomical Society*, Vol. 399, 1106–1118

Moore, B. (1994) Evidence against dissipation-less dark matter from observations of galaxy haloes. *Nature*, Vol. 370, 629–631

Navarro, J. F., Frenk, C. S. & White, S. D. M. (1995) The assembly of galaxies in a hierarchically clustering Universe. *Monthly Notices of the Royal Astronomical Society*, Vol. 275, 56–66

Navarro, J. F., Frenk, C. S. & White, S. D. M. (1996) The Structure of Cold Dark Matter Halos. *The Astrophysical Journal*, Vol. 462, 563–575

Navarro, J. F., Frenk, C. S. & White, S. D. M. (1997) A Universal Density Profile from Hierarchical Clustering. *The Astrophysical Journal*, Vol. 490, 493–508

Opik, E. (1922) An estimate of the distance of the Andromeda Nebula. *The Astrophysical Journal*, Vol. 55, 406–410

eebles, P. J. E. (1980) Nature of the matter distribution now and at Z = 1000. *Physica Scripta*, Vol. 21, 720–724

Peterson, C. J., Rubin, V. C., Ford, W. K. & Roberts, M. S. (1978) Extended rotation curves of high-luminosity spiral galaxies. III - The spiral galaxy NGC 721. *The Astrophysical Journal*, Vol. 226, 770–776

Pizagno, J., Prada, F., Weinberg, D. H., Rix, H.-W., Harbeck, D., Grebel, E. K., Bell, E. F., Brinkmann, J., Holtzman, J. & West, A. (2005) Dark Matter and Stellar Mass in the Luminous Regions of Disk Galaxies. *The Astrophysical Journal*, Vol. 633, 844–856

Pizagno, J., Prada, F., Weinberg, D. H., Rix, H.-W., Pogge, R. W., Grebel, E. K., Harbeck, D., Blanton, M., Brinkmann, J. & Gunn, J. E. (2007) The Tully-Fisher Relation and its Residuals for a Broadly Selected Sample of Galaxies. *The Astrophysical Journal*, Vol. 134, 945–972

Prada, F., Klypin, A. A., Simonneau, E., Betancort-Rijo, J., Patiri, S., Gottlöber, S. & Sanchez-Conde, M. A. (2006) How Far Do They Go? The Outer Structure of Galactic Dark Matter Halos. *The Astrophysical Journal*, Vol. 645, 1001–1011

Quinn, P. J., Salmon, J. K. & Zurek, W. H. (1986) Primordial density fluctuations and the structure of galactic haloes. *Nature*, Vol. 322, 329–335

Regan, M. W. & Vogel, S. N. (1994) The near-infrared structure of M33. *The Astrophysical Journal*, Vol. 434, 536–545

Rubin, V. C. & Ford, W. K. (1970) Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions. *The Astrophysical Journal*, Vol. 159, 379–404

Rubin, V. C., Thonnard, N. & Ford, W. K. (1977) Extended rotation curves of high-luminosity spiral galaxies. I - The angle between the rotation axis of the nucleus and the outer disk of NGC 3672. *The Astrophysical Journal*, Vol. 217, L1–L4

Rubin, V. C., Ford, W. K., Strom, K. M., Strom, S. E. & Romanishin, W. (1978a) Extended rotation curves of high-luminosity spiral galaxies. II - The anemic SA galaxy NGC 4378. *The Astrophysical Journal*, Vol. 224, 782–795

Rubin, V. C., Thonnard, N. & Ford, W. K. (1978b) Extended rotation curves of high-luminosity spiral galaxies. IV - Systematic dynamical properties, Sc through Sc. *The Astrophysical Journal*, Vol. 225, L107–L110

Rubin, V. C.; Ford, W. K. & Roberts, M. S. (1979) Extended rotation curves of high-luminosity spiral galaxies. V - NGC 1961, the most massive spiral known. *The Astrophysical Journal*, Vol. 230, 35–39

Rubin, V. C.; Ford, W. K. & Thonnard, N. (1980) Rotational properties of 21 Sc galaxies with a large range of luminosities and radii, from NGC 4605 (R = 4kpc) to UGC 2885 (R = 122 kpc). *The Astrophysical Journal*, Vol. 238, 471–487

Rubin, V. C., Ford, W. K., Thonnard, N. & Burstein, D. (1982) Rotational properties of 23 Sb galaxies. *The Astrophysical Journal*, Vol. 261, 439–456

Rubin, V. C., Burstein, D., Ford, W. K., & Thonnard, N. (1985) Rotation velocities of 16 Sa galaxies and a comparison of Sa, Sb, and SC rotation properties. *The Astrophysical Journal*, Vol. 289, 81–98

Ryden, B. S. & Gunn, J. E. (1987) Galaxy formation by gravitational collapse. *The Astrophysical Journal*, Vol. 318, 15–31

Salpeter, E. E. (1955) The Luminosity Function and Stellar Evolution. *The Astrophysical Journal*, Vol. 121, 161–167

Seigar, M. S., Barth, A. J. & Bullock, J. S. (2008) A Revised ΛCDM Mass Model for the Andromeda Galaxy. *Monthly Notices of the Royal Astronomical Society*, Vol. 389, 1911–1923

Seigar, M. S. (2011) The Dark Matter Halo Density Profile, Spiral Arm Morphology and Black Hole Mass of M33. *ISRN Astronomy & Astrophysics*, Vol. 2011, 725697

Sellwood, J. A. & McGaugh, S. S. (2005) The Compression of Dark Matter Halos by Baryonic Infall. *The Astrophysical Journal*, Vol. 634, 70–76

Shankar, F., Lapi, A., Salucci, P., de Zotti, G. & Danese, L. (2006) New Relationships between Galaxy Properties and Host Halo Mass, and the Role of Feedbacks in Galaxy Formation. *The Astrophysical Journal*, Vol. 643, 14–25

Simon, J. D., Bolatto, A. D., Leroy, A., Blitz, L. & Gates, E. L. (2005) High-Resolution Measurements of the Halos of Four Dark Matter-Dominated Galaxies: Deviations from a Universal Density Profile. *The Astrophysical Journal*, Vol. 621, 757–776

Spano, M., Marcelin, M., Amram, P., Carignan, C., Epinat, B. & Hernandez, O. (2008) GHASP: an Hα kinematic survey of spiral and irregular galaxies - V. Dark matter distribution in 36 nearby spiral galaxies. *Monthly Notices of the Royal Astronomical Society*, Vol. 383, 297–316

Spergel, D. N. & Steinhardt, P. J. (2000) Observational Evidence for Self-Interacting Cold Dark Matter. *Physical Review Letters*, Vol. 84, 3760-3763

Spergel, D. N., Bean, R., Doré, O., Nolta, M. R., Bennett, C. L., Dunkley, J., Hinshaw, G., Jarosik, N., Komatsu, E., Page, L., Peiris, H. V., Verde, L., Halpern, M., Hill, R. S., Kogut, A., Limon, M., Meyer, S. S., Odegard, N., Tucker, G. S., Weiland, J. L. & Wright, E. L. (2007) Three-Year Wilkinson Anisotropy Probe (WMAP) Observations: Implications for Cosmology. *The Astrophysical Journal Supplement*, Vol. 170, 377–408

Springel, V., White, S. D. M., Jenkins, A., Frenk, C. S., Yoshida, N., Gao, L., Navarro, J., Thacker, R., Croton, D., Helly, J., Peacock, J. A., Cole, S., Thomas, P., Couchman, H.,

Evrard, A., Colberg, J., & Pearce, F. (2005) Simulations of the Formation, Evolution and Clustering of Galaxies and Quasars. *Nature*, Vol. 435, 629–636

Strom, S. E., Strom, K. M., Wells, D. C., Forte, J. C., Smith, M. G. & Harris, W. E. (1981) The halo globular clusters of the giant elliptical galaxy Messier 87. *The Astrophysical Journal*, Vol. 245, 416–453

Swaters, R. A., Madore, B. F., van den Bosch, F. C. & Balcells, M. (2003) The Central Mass Distribution in Dwarf and Low Surface Brightness Galaxies. *The Astrophysical Journal*, Vol. 583, 732–751

Tamm, A., Tempel, E. & Tenjes, P. (2007) Visible and dark matter in M31 - I. Properties of stellar components. (arXiv:0707.4375)

Tempel, E., Tamm, A. & Tenjes, P. (2007) Visible and dark matter in M 31 - II. A dynamical model and dark matter density distribution. (arXiv:0707:4374)

Tully, R. B. & Fisher, J. R. (1977) A new method of determining distances to galaxies. *Astronomy & Astrophysics*, Vol. 54, 661–673

Valenzuela, O., Rhee, G., Klypin, A., Governato, F., Stinson, G., Quinn, T., & Wadsley, James (2007) Is There Evidence for Flat Cores in the Halos of Dwarf Galaxies? The Case of NGC 3109 and NGC 6822. *The Astrophysical Journal*, Vol. 657, 773–789

Warren, M. S., Quinn, P. J., Salmon, J. K. & Zurek, W. H. (1992) Dark halos formed via dissipationless collapse. I - Shapes and alignment of angular momentum. *The Astrophysical Journal*, Vol. 399, 405–425

Wechsler, R. H., Bullock, J. S., Primack, J. R., Kravtsov, A. V. & Dekel, A. (2002) Concentrations of Dark Halos from Their Assembly Histories. *The Astrophysical Journal*, Vol. 568, 52–70

Weinberg, D. H., Colombi, S., Davé, R. & Katz, N. (2008) Baryon Dynamics, Dark Matter Substructure, and Galaxies. *The Astrophysical Journal*, Vol. 678, 6–21

West, M. J., Dekel, A. & Oemler, A. (1987) Profiles of clusters of galaxies - Cosmological scenarios versus observations. *The Astrophysical Journal*, Vol. 316, 1–22

White, S. D. M. & Rees, M. J. (1978) Core condensation in heavy halos - A two-stage theory for galaxy formation and clustering. *Monthly Notices of the Royal Astronomical Society*, Vol. 183, 341–358

White, S. D. M. & Frenk, C. S. (1991) Galaxy formation through hierarchical clustering. *The Astrophysical Journal*, Vol. 379, 52–79

Yang, X., Mo, H. J. & van den Bosch, F. C. (2003) Constraining galaxy formation and cosmology with the conditional luminosity function of galaxies. *Monthly Notices of the Royal Astronomical Society*, Vol. 339, 1057–1080

Zappacosta, L., Buote, D. A., Gastaldello, F., Humphrey, P. J., Bullock, J., Brighenti, F. & Mathews, W. (2006) The Absence of Adiabatic Contraction of the Radial Dark Matter Profile in the Galaxy Cluster A2589. *The Astrophysical Journal*, Vol. 650, 777–790

Zentner, A. R. & Bullock, J. S. (2002) Inflation, cold dark matter, and the central density problem. *Physical Review D*, Vol. 66, 043003

Zurek, W. H.; Quinn, P. J.; Salmon, J. K. (1988) Rotation of halos in open and closed universes - Differentiated merging and natural selection of galaxy types. *The Astrophysical Journal*, Vol. 330, 519–534

Part 2

Dark Energy

Holographic Dark Energy Model with Chaplygin Gas

B. C. Paul

Physics Department, North Bengal University,
Siliguri, Dist. : Darjeeling, Pin : West Bengal, India
India

1. Introduction

Recent cosmological observations namely, high redshift surveys of type Ia Supernovae, CMB etc. predict that the present universe is passing through an accelerating phase of expansion [1]. The discovery of anisotropy in the Cosmic Microwave Background (CMB) by the Cosmic Microwave Background Explorer (in short, COBE) satellite marked a major advance in cosmology. CMB anisotropy in general alongwith COBE data now provide some of the most powerful tests of cosmological theories, particularly theories concerning the formation of large-scale structure of the universe. It is now generally believed that the universe has emerged out of an inflationary phase in the past. It has been recently predicted that the present universe is undergoing an accelerating phse of expansion. The above two observational facts in the universe do not find explanation in the framework of Einstein general theory of Relativity (GTR) with normal matter. It is known that early inflation may be realized in a semiclassical theory of gravity where matter is described by quantum fields [2]. Starobinsky obtained inflationary solution considering a curvature squared term in the Einstein-Hilbert action [3] long before the advent of inflation was known. However, the efficacy of inflation is known only after the seminal work of Guth who first employed the phase transition mechanism to accommodate inflation. Thereafter more than a dozen of inflationary model came up. The early inflationary universe can be realized either by (i) a modification of gravitational sector of action introducung higher order terms or [4], (ii) a modification of the matter sector instead of perfect fluid, matter is described by quantum fields. Another way of realizing inflation is to consider imperfect fluid. In the case of matter, the equation of state (in short, EOS) is $p = \omega \rho$ where p represents the pressure and ρ represents the energy density, inflation is permitted if $\omega = -1$. For a homogeneous scalar field the condition for inflation is achieved when the potential energy of the field dominates over the kinetic energy. The inflationary period ends when the scalar field or the inflaton field reaches the minimum of the potential. Thereafter the field oscillates and a small comoving volume grows to accommodate the whole universe. The universe decelerates during matter dominated era but the recent observation is that the universe is accelerating. The present accelerating phase can be realized with EOS parameter $\omega < -1$. The usual fields in the standard model of particle physics are not suitable to obtain such accelerating phase of the present universe. In fact it is a challenge

to theoretical physics to describe the origin of such matter density. Recent astronomical data when interpreted in the context of the big-bang model have provided some interesting information about the composition of the universe. The total energy density is shared among a number of fluids that comprises the universe. While Big-Bang nucleosynthesis data suggest that baryonic matter can account for only about 4 % of the total energy density, the cold dark matter (CDM) is about 23 % and the third part, called dark energy, constitutes the remaining 73 % of the total energy density of the universe. To accommodate such a huge energy density various kinds of exotic matters are taken into account as a possible candidate for the dark energy. Among the different theories put forwarded in the literature in recent times, the single component fluid known as Chaplygin gas with an equation of state (EOS) $p = -\frac{B}{\rho}$ [5], where ρ and p are the energy density and pressure respectively and B is a constant has attracted large interest in cosmology. The above equation of state, however, has been conceived in studies of adiabatic fields. It was used to describe lifting forces on a plane wing in aerodynamics process. In cosmology, although it admits an accelerating universe, fails to address structure formation and cosmological perturbation power spectrum [11]. Subsequently, a modified form of the equation of state $p = -\frac{B}{\rho^\alpha}$ with $0 \le \alpha \le 1$ was also considered to construct a viable cosmological model [6, 7], which is known as generalized Chaplygin gas (GCG). It has two free parameters B and α respectively. The fluid behaves initially like dust for small size of the universe, but at a later epoch the fluid may be described by an equation of state $p = \omega\rho$. It has string connection, the above EOS can be obtained from the Nambu-Goto action for a D-brane moving in a (D+2) -dimensional space-time in the light cone parametrization. The EOS for GCG has been further modified to

$$p = A\rho - \frac{B}{\rho^\alpha} \quad with \ 0 \le \alpha \le 1. \tag{1}$$

where A is an equation of state parameter and B is a constant, known as modified GCG. The modified GCG has three free parameters. In the early universe when the scale factor of the universe $a(t)$ was small, the modified GCG gas corresponds to a barotropic fluid (if one considers $A = \frac{1}{3}$ it corresponds to radiation and $A = 0$ it corresponds to matter). Thus at one extreme end modified GCG behaves as ordinary fluid and at the other extreme when the universe is sufficiently large it behaves as cosmological constant which can be fitted to a ΛCDM model. In a flat Friedmann model it is shown [6] that the modified generalized Chaplygin gas may be equivalently described in terms of a homogeneous minimally coupled scalar field ϕ. Barrow [10] has outlined a method to fit Chaplygin gas in FRW universe. Gorini et al. [11] using the above scheme obtained the corresponding homogeneous scalar field $\phi(t)$ in a potential $V(\phi)$ which can be used to obtain a viable cosmological model with modified Chaplygin gas. Another form of EOS for Chaplygin gas [12] is considered recently which is given by

$$p = A\rho - \frac{B(a)}{\rho^\alpha} \quad with \ 0 \le \alpha \le 1, \tag{2}$$

with a variable $B = B_0 a^{-3n}$, B_0 is a constant and a is the scale factor of the universe. The fluid described by the above equation is of much importance now-a-days which is known as variable Chaplygin gas. Using the scheme given by Gorini and his coworkers the corresponding homogeneous scalar field $\phi(t)$ and its potential $V(\phi)$ may be obtained. Guo and Zhang [13] obtained cosmological model using the EOS for variable Chaplygin gas. It is

therefore important to track the dark energy component of the energy density in terms of the above EOS. Cosmological observational results will be used to obtain the necessary constraints [14].

Recently holographic principle [15, 16] is incorporated in cosmology [17-20] to track the dark energy content of the universe following the work of Cohen *et al.* [21]. Holographic principle is a speculative conjecture about quantum gravity theories proposed by G't Hooft. The idea is subsequently promoted by Fischler and Susskind [15] claiming that all the information contained in a spatial volume may be represented by a theory that lives on the boundary of that space. For a given finite region of space it may contain matter and energy within it. If this energy be less than a critical value then the region collapses to a black hole. A black hole is known theoretically to have an entropy which is proportional to its surface area of its event horizon. The event horizon of a black hole encloses a volume, thus a more massive black hole have larger event horizon and encloses larger volume. The most massive black hole that can be fitted in a given region is the one whose event horizon corresponds exactly to the boundary of the given region under consideration. The maximal limit of entropy for an ordinary region of space is directly proportional to the surface area of the region and not to its volume. Thus, according to holographic principle, under suitable conditions all the information about a physical system inside a spatial region is encoded in the boundary. The basic idea of a holographic dark energy in cosmology is that the saturation of the entropy bound may be related to an unknown ultra-violet (UV) scale Λ to some known cosmological scale in order to enable it to find a viable formula for the dark energy which may be quantum gravity in origin and it is characterized by Λ. The choice of UV-Infra Red (IR) connection from the covariant entropy bound leads to a universe dominated by blackhole states. According to Cohen *et al.* [21] for any state in the Hilbert space with energy E, the corresponding Schwarzschild radius $R_s \sim E$, may be less than the IR cut off value L (where L is a cosmological scale). It is possible to derive a relation between the UV cutoff $\rho_\Lambda^{1/4}$ and the IR cutoff which eventually leads to a constraint $\left(\frac{8\pi G}{c^2}\right) L^3 \left(\frac{\rho_\Lambda}{3}\right) \leq L$ [22] where ρ_Λ is the energy density corresponding to dark energy characterized by Λ, G is Newton's gravitational constant and c is a parameter in the theory. The holographic dark energy density is

$$\rho_\Lambda = 3c^2 M_P^2 L^{-2}, \qquad (3)$$

where $M_P^{-2} = 8\pi G$. It is known that the present acceleration may be described if $\omega_\Lambda = \frac{p_\Lambda}{\rho_\Lambda} < -\frac{1}{3}$. If one considers $L \sim \frac{1}{H}$ it gives $\omega_\Lambda = 0$. A holographic cosmological constant model based on Hubble scale as IR cut off does not permit accelerating universe. It is also examined [17] that the holographic dark energy model based on the particle horizon as the IR cutoff is not suitable for an accelerating universe. However, an alternative model of dark energy using particle horizon in closed model is also proposed [23]. Li [18] has obtained an accelerating universe considering event horizon as the cosmological scale. The model is consistent with the cosmological observations. Thus to have a model consistent with observed universe one should adopt the covariant entropy bound and choose L to be event horizon [24]. Considering a correspondence of holographic dark energy and Chaplygin gas the field potential is reconstructed [25, 26]. In this chapter we consider variable Chaplygin gas whose EOS is given by eq. (2) and set up a correspondence with holographic dark energy to reconstruct scalar field potential. The holographic description of the variable Chaplygin

gas as a dark energy in FRW universe is considered to explain the dark energy needed for an accelerating universe at late epoch.

The chapter is organized as follows : in sec. 2, the relevant field equation with modified variable Chaplygin gas in FRW universe is presented. Considering correspondence of holographic dark energy fields with modified variable chaplygin gas, we determine the field and the corresponding potential is reconstructed. in sec. 3, squared speed of sound for holographic dark energy is evaluated for stability analysis. Finally in sec. 4, a brief discussion is given.

2. Field equation and modified variable chaplygin gas

The Einstein's field equation is given by

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi G\, T_{\mu\nu} \tag{4}$$

where $T_{\mu\nu}$ is the energy momentum tensor.

We consider a Robertson-Walker (RW) metric given by

$$ds^2 = -dt^2 + a^2(t)\left[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + sin^2\theta\, d\phi^2)\right] \tag{5}$$

where $a(t)$ is the scale factor of the universe. The energy momentum tensor is $T_\nu^\mu = (\rho, -p, -p, -p)$ where ρ and p are energy density and pressure respectively.

Using RW metric given by (5) and the energy momentum tensor, the Einstein's field equation (4) yields

$$H^2 + \frac{k}{a^2} = \frac{1}{3M_P^2}\rho \tag{6}$$

where we use $8\pi G = M_P^2$, $H = \frac{\dot{a}}{a}$ is the Hubble parameter. The conservation equation for matter is given by

$$\frac{d\rho}{dt} + 3H(\rho + p) = 0 \tag{7}$$

where $\rho = \rho_{matter} + \rho_\Lambda$, where ρ_Λ is the energy density corresponding to Λ. For modified variable chaplygin gas (henceforth, VCG), we use EOS given by eq. (2) in eq. (6), which yields

$$\rho_\Lambda = \left(\frac{(1+\alpha)B_o}{(1+\alpha)(1+A)-n}\frac{1}{a^{3n}} + \frac{C_o}{a^m}\right)^{\frac{1}{1+\alpha}} \tag{8}$$

where C_o is an integration constant and we denote $m = 3(1+A)(1+\alpha)$. However, for modified Chaplygin gas (i.e., for $n = 0$) it reduces to

$$\rho_\Lambda = \left(\frac{B_o}{1+A} + \frac{C_o}{a^m}\right)^{\frac{1}{1+\alpha}}.$$

We now define the following

$$\Omega_\Lambda = \frac{\rho_\Lambda}{\rho_{cr}},\ \Omega_m = \frac{\rho_m}{\rho_{cr}},\ \Omega_k = \frac{k}{a^2H^2} \tag{9}$$

where $\rho_{cr} = 3M_P^2 H^2, \Omega_\Lambda, \Omega_m$ and Ω_k represent density parameter corresponding to Λ, matter and curvature respectively.

We assume here that the origin of dark energy is a scalar field. Making use of Barrow's scheme [10], we get the following

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V(\phi) = \rho = \left(\frac{B_1}{a^{3n}} + \frac{C_o}{a^m}\right)^{\frac{1}{\alpha+1}}, \tag{10}$$

$$p_\phi = \frac{1}{2}\dot{\phi}^2 - V(\phi) = p = \frac{\frac{(n-1-\alpha)B_1}{(1+\alpha)a^{3n}} + \frac{AC_o}{a^m}}{\left(\frac{B_1}{a^{3n}} + \frac{C_o}{a^m}\right)^{\frac{\alpha}{1+\alpha}}}, \tag{11}$$

where $B_1 = \frac{(1+\alpha)B_O}{(1+\alpha)(1+A)-n}$. Now the corresponding scalar field potential and its kinetic energy term is obtained from above which are given by

$$V(\phi) = \frac{\frac{C_o(1-A)}{2a^m} + \frac{1+\alpha-n}{1+\alpha}\frac{B_1}{a^{3n}}}{\left(\frac{B_1}{a^{3n}} + \frac{C_o}{a^m}\right)^{\frac{\alpha}{1+\alpha}}}, \tag{12}$$

$$\dot{\phi}^2 = \frac{\frac{nB_1}{(1+\alpha)a^{3n}} + \frac{C_o(1+A)}{a^m}}{\left(\frac{B_1}{a^{3n}} + \frac{C_o(1+A)}{a^m}\right)^{\frac{\alpha}{1+\alpha}}} \tag{13}$$

For MCG in a flat universe($k = 0$), eq. (13) can be integrated which yields

$$\phi = \pm\frac{2}{\sqrt{m}} \sinh^{-1}\left[\sqrt{\frac{C_o}{B_1}} a^{-\frac{m}{2}}\right] \tag{14}$$

when $n = 0$ [27] and the corresponding potential is given by

$$V(\phi) = \frac{\frac{B_o}{1+A} + \frac{B_o(1-A)}{2(1+A)}Sinh^2\left(\sqrt{\frac{3(1+A)(1+\alpha)}{2}}\phi\right)}{\left(\frac{B_o}{1+A}\right)^{\frac{\alpha}{1+\alpha}} cosh^{\frac{2\alpha}{1+\alpha}}\left(\sqrt{\frac{3(1+A)(1+\alpha)}{2}}\phi\right)} \tag{15}$$

The above equation further reduces to that obtained in *Ref.* [26] for $\alpha = 1$, $A = 0$ and $n = 0$. Now we consider that the scalar field model of dark energy described by modified variable Chaplygin gas which corresponds to holographic dark energy of the universe. In this section we reconstruct the correponding potential. Let us now consider a non-flat universe with $k \neq 0$ and use the holographic dark energy density given in (3) as

$$\rho_\Lambda = 3c^2 M_P^2 L^{-2}, \tag{16}$$

where L is the cosmological length scale for tracking the field corresponding to holographic dark energy in the universe. The parameter L is defined as

$$L = ar(t). \tag{17}$$

where $a(t)$ is the scale factor of the universe and $r(t)$ is relevant to the future event horizon of the universe. Using Robertson-Walker metric one gets [19]

$$L = \frac{a(t)}{\sqrt{|k|}} \ sin \ \left[\sqrt{|k|}R_h(t)/a(t) \right] \quad for \quad k = +1,$$

$$= R_h \ for \ k = 0,$$

$$= \frac{a(t)}{\sqrt{|k|}} \ sinh \ \left[\sqrt{|k|}R_h(t)/a(t) \right] \quad for \quad k = -1. \tag{18}$$

where R_h represents the event horizon which is given by

$$R_h = a(t) \int_t^\infty \frac{dt'}{a(t')} = a(t) \int_o^{r_1} \frac{dr}{\sqrt{1 - kr^2}}. \tag{19}$$

Here R_h is measured in r direction and L represents the radius of the event horizon measured on the sphere of the horizon. Using the definition of $\Omega_\Lambda = \frac{\rho_\Lambda}{\rho_{cr}}$ and $\rho_{cr} = 3M_P^2H^2$, one can derive [20]

$$HL = \frac{c}{\sqrt{\Omega_\Lambda}}. \tag{20}$$

Using eqs. (19)- (20), we determine the rate of change of L with respect to t which is

$$\dot{L} = \frac{c}{\sqrt{\Omega_\Lambda}} - \frac{1}{\sqrt{|k|}} \ cos \ \left(\frac{\sqrt{|k|} \ R_h}{a(t)} \right) \quad for \quad k = +1,$$

$$= \frac{c}{\sqrt{\Omega_\Lambda}} - 1 \ for \ k = 0,$$

$$= \frac{c}{\sqrt{\Omega_\Lambda}} - \frac{1}{\sqrt{|k|}} \ cosh \ \left(\frac{\sqrt{|k|} \ R_h}{a(t)} \right) \quad for \quad k = -1. \tag{21}$$

Using eqs. (16) -(21) , we obtain the holographic energy density ρ_Λ, which is given by

$$\frac{d\rho_\Lambda}{dt} = -2H \left[1 - \frac{\sqrt{\Omega_\Lambda}}{c} \frac{1}{\sqrt{|k|}} f(X) \right] \rho_\Lambda, \tag{22}$$

here we use the notation, henceforth,

$$f(X) = cos(X) \ [1, cosh(X)] \ for \ k = 1 \ [0, -1], \tag{23}$$

with $X = \frac{R_h}{a(t)}$. The energy conservation equation is

$$\frac{d\rho_\Lambda}{dt} + 3H(1 + \omega_\Lambda)\rho_\Lambda = 0 \tag{24}$$

which is used to determine the equation of state parameter

$$\omega_\Lambda = - \left(\frac{1}{3} + \frac{2\sqrt{\Omega_\Lambda}}{3c} f(X) \right). \tag{25}$$

Now we assume holographic dark energy density which is equivalent to the modified variable Chaplygin gas energy density. The corresponding energy density may be obtained using eq. (8). The EOS parameter follows from eq. (2)

$$\omega = \frac{p}{\rho} = A - \frac{B(a)}{\rho^{\alpha+1}}.$$

(26)

We now consider correspondence between the holographic dark energy and modified Chaplygin gas energy density. Using eqs. (8), (16) and (22), one obtains

$$B_o = (3c^2 M_P^2 L^{-2})^{\alpha+1} \, a^{3n} \left[A + \frac{1}{3} + \frac{2\sqrt{\Omega_\Lambda}}{3c} f(X) \right],$$

(27)

$$C_o = (3c^2 M_P^2 L^{-2})^{\alpha+1} \, a^{m} \left[1 - \frac{(1+\alpha)(A + \frac{1}{3} + 2\frac{\sqrt{\Omega_\Lambda}}{3c} f(X))}{(1+\alpha)(A+1) - n} \right].$$

(28)

Consequently one determines the scalar field potential which is given by

$$V(\phi) = \frac{3c^2 M_P^2 L^{-2}}{2} \left[1 - A + \frac{(1+\alpha)(1+A) - 2n}{(1+\alpha)(1+A) - n} \left(A + \frac{1}{3} + \frac{2\sqrt{\Omega_\Lambda}}{3} f(X) \right) \right],$$

(29)

and the corresponding kinetic energy of the field is

$$\dot{\phi}^2 = 2c^2 M_P^2 L^{-2} \left[1 - \frac{\sqrt{\Omega_\Lambda}}{c} f(X) \right].$$

(30)

It is interesting to note that for $n = 0$ the potential reduces to the form that obtained by Paul *et al.* [27] and for $n = 0$ and $A = 0$, it reduces to that form obtained by Setare [25] (where B_o is to be replaced by A). We now substitute $x \, (= \ln a(t))$, to transform the time derivative to the derivative with logarithm of the scale factor, which is the most useful function in this case. Consequently from eq. (30) one obtains

$$\phi' = M_P \sqrt{2\Omega_\Lambda \left(1 - \frac{\sqrt{\Omega_\Lambda}}{c} f(X) \right)}$$

(31)

where $()'$ prime represents derivative with respect to x. Thus, the evolution of the scalar field is given by

$$\phi(a) - \phi(a_o) = \sqrt{2} M_P \int_{\ln a_o}^{\ln a} \sqrt{\Omega_\Lambda \left(1 - \frac{\sqrt{\Omega_\Lambda}}{c} f(X) \right)} \, dx.$$

(32)

3. Squared speed for Holographic Dark Energy (HDE) or stability of HDE :

We consider a closed universe model ($k = 1$) in this case. The dark energy equation of state parameter given by eq. (32) reduces to

$$\omega_\Lambda = -\frac{1}{3} \left(1 + \frac{2}{c} \sqrt{\Omega_\Lambda} \cos y \right)$$

(33)

where $y = \frac{R_H}{a(t)}$. The minimum value it can take is $\omega_{min} = -\frac{1}{3}\left(1 + 2\sqrt{\Omega_\Lambda}\right)$ and one obtains a lower bound $\omega_{min} = -0.9154$ for $\Omega_\Lambda = 0.76$ with $c = 1$. Taking variation of the state parameter with respect to $x = \ln a$, we get [17]

$$\frac{\Omega'_\Lambda}{\Omega^2_\Lambda} = (1 - \Omega_\Lambda)\left(\frac{2}{c}\frac{1}{\Omega_\Lambda}cos\, y + \frac{1}{1 - a\gamma}\frac{1}{\Omega_\Lambda}\right) \tag{34}$$

and the variation of equation of state parameter becomes

$$\omega'_\Lambda = -\frac{\sqrt{\Omega_\Lambda}}{3c}\left[\frac{1 - \Omega_\Lambda}{1 - \gamma a} + \frac{2\sqrt{\Omega_\Lambda}}{c}\left(1 - \Omega_\Lambda cos^2 y\right)\right], \tag{35}$$

where $\gamma = \frac{\Omega^o_k}{\Omega^o_m}$. We now introduce the squared speed of holographic dark energy fluid as

$$v^2_\Lambda = \frac{dp_\Lambda}{d\rho_\Lambda} = \frac{\dot{p}_\Lambda}{\dot{\rho}_\Lambda} = \frac{p'_\Lambda}{\rho'_\Lambda}, \tag{36}$$

where varaiation of eq. (30) w.r.t. x is given by

$$p'_\Lambda = \omega'_\Lambda \rho_\Lambda + \omega_\Lambda \rho'_\Lambda. \tag{37}$$

Using the eqs. (36) and (37) we get

$$v^2_\Lambda = \omega'_\Lambda \frac{\rho_\Lambda}{\rho'_\Lambda} + \omega_\Lambda$$

which now becomes

$$v^2_\Lambda = -\frac{1}{3} - \frac{2}{3c}\sqrt{\Omega_\Lambda}\, cosy + \frac{1}{6c}\sqrt{\Omega_\Lambda}\left[\frac{\frac{1 - \Omega_\Lambda}{1 - \gamma a} + \frac{2}{c}\sqrt{\Omega_\Lambda}\left(1 - \Omega_\Lambda\, cos^2 y\right)}{1 - \frac{\Omega_\Lambda}{c}\, cosy}\right]. \tag{38}$$

The variation of v^2_Λ with Ω_Λ is shown in fig. 1 for different y values. It is found that for a given value of c, a, γ, the model admits a positive squared speed for $\Omega_\Lambda > 0$. However, Ω_Λ is bounded below otherwise instability develops. We note also that for $\frac{(2n+1)\pi}{2} < y < \frac{(2n+3)\pi}{2}$, (where n is an integer) no instability develops. We plot the case for $n = 0$ in fig. 2, it is evident that for $y \le \frac{\pi}{2}$ and $y \ge \frac{3\pi}{2}$, the squared speed for holographic dark energy becomes negative which led to instability. But for the region $\frac{\pi}{2} < y < \frac{3\pi}{2}$ with $n = 0$ no such instability develops. It is also found that for $y = 0$ i.e., in flat case the holographic dark energy model is always unstable [28].

4. Discussions

Holographic dark energy model is explored here in a FRW universe with a scalar field which describes the modified variable Chaplygin gas. We consider correspondence of holographic dark energy and the modified variable Chaplygin gas to reconstruct the potential. Since a complete theory of quantum gravity is yet to emerge, we adopt the above approach to understand the nature of dark energy. We determine the evolution of the field and reconstruct the potential of the Holographic dark energy in the case of flat, closed and open models of
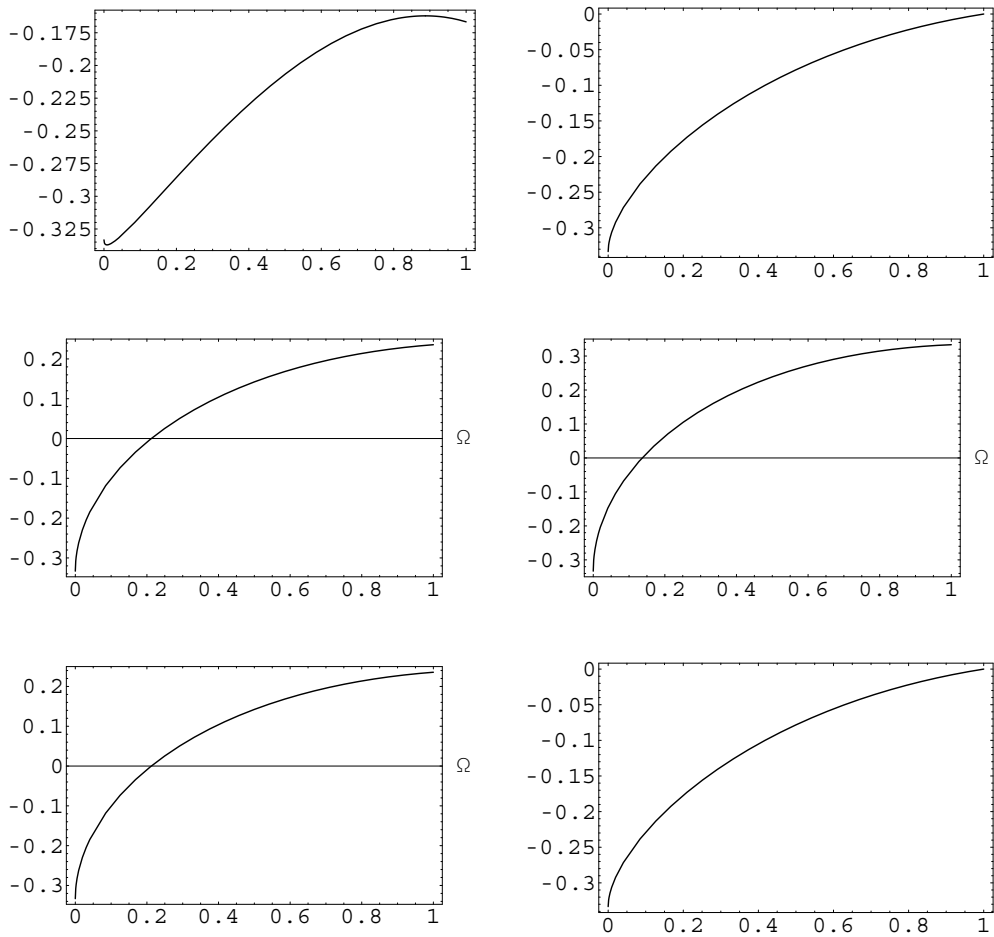
Fig. 1. shows the plot of v_Λ^2 versus Ω_Λ for different values of y with $c = 1$, $\gamma = 1/3$ and $a = 1$, in the first array the figures are for $y = \frac{\pi}{3}$ and $y = \frac{\pi}{2}$, in the second array for $y = \frac{1.5\,\pi}{2}$, $y = \pi$ and in the third array for $y = \frac{2.5\,\pi}{2}$, $y = \frac{3\pi}{2}$.

the universe. Although the cosmological observations predict a flat model of the universe, a closed universe with small positive curvature ($\Omega_k = 0.01$) is compatible with observations. So, in this paper we considered non-flat case also. We give here a generalized expression for the potential and the kinetic energy term considering a modified variable Chaplygin gas [12, 13]. The holographic dark energy field and the corresponding potential depend on three parameters namely, A, α and n. The potential and the kinetic energy given by eqs. (11) and (12) reduce to that form obtained by Setare [26] for $A = 0$, $n = 0$ and $\alpha = 1$. However, the result obtained by Paul *et al.* [27] recovered for $n = 0$. The stability of the holographic dark energy is studied in sec. 3 and found that the stability depends on the parameter Ω_Λ. The evolution of the holograpic dark energy field follows same pattern in the modified Chaplygin gas, generalized Chaplygin gas and in the variable Chaplygin gas. However, the field potential is found to differ which, however, depends on the EOS parameters $A = 0$, $n = 0$ and $\alpha = 1$. It is found that holographic field potential $V(\phi)$ picks up a constant value for $n = \frac{(1+\alpha)(1+A)}{2}$. It is noted that the model permits a free holographic field when one chooses $A = 1$ and $n = 1 + \alpha$.

5. Acknowledgement

6. References

[1] A. G Riess *et al.*, *Astrophys. J.* 607, 665 (2004); S. Perlmutter *et al., Nature* 51, 391 (1998); S. Perlmutter *et al., Astrophys. J.* 598, 102 (2003); S. Perlmutter *et al., Astrophys. J.* 517, 565 (1999);G. F. Smoot *et. al., Ap. J.* 369 L1 (1992); P. de Bernardis *et al. Nature* 404 955 (2000); M. Tegmark *et al., Phys. Rev* D 69 , 103501 (2004); D. N. Spergel *et al.*, arXive ; astro-ph/0603449.

[2] A. H. Guth, *Phys. Rev.* D 23, 347 (1981); A. D. Linde, *Phys. Lett.* B 108, 389 (1982); *Rep. Prog. Phys.* 47, 925 (1984); R. H. Brandenberger, *Rev. Mod. Phys.* 57, 1 (1985); D. H. Lyth, D. Roberts and M Smith, *Phys. Rev.* D 57, 7120 (1998); L. Mersini, *Mod. Phys. Lett.* A 16, 1933 (2001); A. Linde, *Int. J. Mod. Phys.* A17, 89 (2002).

[3] A. A. Starobinsky, *Phys. Lett.* B 99, 24 (1980).

[4] R. Fabri and M. D. Pollock, *Phys. Lett.* B 125, 445 (1983); A. A. Starobinsky, *Sov. Astron. Lett.* 9, 302 (1983), *JETP Lett.* 42, 152 (1986); S. W. Hawking and J. C. Luttrell, *Nucl. Phys.* B 247, 250 (1984); L. A. Kofmann, A. D. Linde and A. A. Satrobinsky, *Phys. Lett.* B 157, 361 (1985); A. Vilenkin, *Phys. Rev.* D 32, 2511 (1985); M. B, Mijic, M. S. Morris and W. Suen, *Phys. Rev.* D 34, 2934 (1986); S. Gottlober and V. Muller, *Class. Quantum Grav.* 3, 183 (1986); B. C. Paul, D. P. Datta and S. Mukherjee *Mod. Phys. Lett.* A 3, 843 (1988); G. Magnano and S. M. Sokolowski, *Phys. Rev.* D 50, 5039 (1994); B. C. Paul and A. Saha, *Int. J. Mod. Phys.* D 11, 493 (2002); S. Mukherjee, B. C. Paul, A. Beesham and S. D. Maharaj, arXive: gr-qc/0505103; P.S. Debnath and B. C. Paul, *Int. J. Mod. Phys.* D 15, 189 (2006);

S. Mukherjee, B. C. Paul, N. K. Dadhich, A. Beesham and S. D. Maharaj , *Class. Quantum Grav.* 23, 6927 (2006).

[5] A. Y. Kamenshchik, U. Moschella and V. Pasquier, *Phys. Lett.* B 511, 265 (2001); V. Gorini, A. Kamenshchik, U. Moschella and V. Pasquier, arXive: gr-qc/0403062.

[6] V. Gorini, A. Kamenshchik and U. Moschella, *Phys. Rev.* D 67 063509 (2003); U. Alam, V. Sahni, T. D. Saini and A. A. Starobinsky, *Mon. Not. Roy. Astron. Soc.* 344, 1057 (2003)

[7] M. C. Bento, O. Bertolami and A. A. Sen, *Phys. Rev.* D 66, 043507 (2002); V. Sahni, T. D. Saini, A. A. Starobinsky and U. Alam, *JETP Lett.* 77, 201 (2003).

[8] H. B. Benaoum, *Accelerated Universe from Modified Chaplygin gas and tachyonic fluid*, arXive : hep-th/0205140.

[9] S. Nojiri and S. D. Odintsov, *Phys. Rev.* D 70, 103522 (2004), arXive: hep-th/0408170; *Phys. Rev.* D 72, 023003 (2005), arXive: hep-th/0505215.

[10] J. D. Barrow, *Nucl. Phys.* B 310, 743 (1988), *Phys. Lett.* B 235, 40 (1990).

[11] V. Gorini, A. Kamenshchik, U. Moschella and V. Pasquier, *Phys. Rev.* D 69 123512 (2004), A. Y. Kamenshchik, U. Moschella and V. Pasquier, *Phys. Lett.* B 511 265 (2001).

[12] M. Jamil and M. A. Rashid, *Europhys. J* C 58, 111 (2008).

[13] Z-K. Guo and Y-Z. Zhang, *Phys. Lett.* B 645, 326 (2007).

[14] G. Sethi, S. K. Singh, P. Kumar, D. Jain and A. Dev, *Int. J. Mod. Phys.* D 15, 1089 (2006); Z-K. Guo and Y-Z. Zhang, arXive: astro-ph/0509790.

[15] W. Fischler and L. Susskind, *Holography and Cosmology* arXive : hep-th/9806039; L. Susskind, *Holography in the flat space limit*, arXive : hep-th/9901079; D. Bigatti and L. Susskind, *TASI Lectures on the Holographic Principle*, arXive : hep-th/0002044.

[16] R. Bousso, *JHEP* 9907, 004 (1999), *JHEP* 9906, 028 ((1999); *Class. Quantum Grav,* 17, 997 (2000).

[17] S. D. H. Hsu, *Phys. Lett.* B 594, 13 (2004).

[18] M. Li, *Phys. Lett.* B 603, 1 (2004).

[19] J. Zhang, X. Zhang and H. Liu, *Phys. Lett.* B 651, 84 (2007), arXive: 0706.1185; X. Zhang, *Phys. Lett.* B 648, 1 (2007); M. R. Setare, *Phys. Lett.* B 653, 116 (2007); N. Banerjee and D. Pavon, *Phys. Lett.* B 647, 477 (2007); M. R. Setare, *JCAP,* 0701, 023 (2007); B. Chen, M. Li and Y. Wang, *Nucl. Phys.* B 774, 256 (2007); M. R. Setare, *Phys. Lett.* B 644, 99 (2007); M. R. Setare, *Eur. Phys. J.* C 50 991 (2007); M. R. Setare and S. Shafei, *JCAP* 0609, 011 (2006); M. R. Setare, *Phys. Lett.* B 642, 1 (2006); J. P. Beltran Almeida and J. G. Pereirs, *Phys. Lett.* B 636, 75 (2006), arXive : gr-qc/0602103; X. Zhang, *Phys. Rev.* 74, 103505 (2006), arXive : astro-ph/0609699; S. Nojiri and S. D. Odintsov, *Gen. Rel. Grav.* 38, 1285 (2006), hep-th/0506212; Y. Gong and Y.Z. Zhang, *Class. Quantum Grav.* 22, 4895 (2005), hep-th/0505175; X. Zhang and F. Q. Wu, *Phys. Rev.* D 72, 043524 (2005); D. Pavon and W. Zimdahl, *AIP Conf. Proce.* 841, 356 (2006), arXive: hep-th/0511053.

[20] Q. G. Huang and M. Li, *JCAP* 8, 13 (2004), arXive: astro-ph/0404229.

[21] A. G. Cohen, D. B. Kaplan and A. E. Nelson, *Phys. Rev. Lett.* 82, 4971 (1999).

[22] P. Horava and D. Minic, *Phys. Rev. Lett.* 85, 1610 (2000); S. Thomas, *Phys. Rev. Lett.* 89, 081301 (2002).

[23] F. Simpson, *JCAP* 0703, 016 (2007).

[24] F. Leblond and A. W. Peet, *JHEP*, 0304, 048 (2003); N. Lambert, H. Liu and J. Maldacena, *JCAP* 0703, 014 (2007) arXive :hep-th/0303139; C. J. Km. H. B. Kim. Y. B. Kim and O. K. Kwon, *JHEP*, 0303, 008 (2003); D. A. Steer and F. Vernizzi, *Phys. Rev. ,* D 70 043527 (2004).

[25] M. R. Setare, *Phys. Lett.* B 654, 1 (2007).

[26] M. R. Setare, *Phys. Lett.* B 648, 329 (2007).

[27] B. C. Paul, P. S. Thakur and S. Ghose, arXive:0809.3491.

[28] Y. S. Myung, *Phys. Lett.* B 652, 223 (2007).

Strong Lensing Systems as Probes of Dark Energy Models and Non-Standard Theories of Gravity

Marek Biesiada

Institute of Physics, Department of Astrophysics and Cosmology,
University of Silesia
Poland

1. Introduction

Gravitational lensing is a consequence of light deflection in gravitational field. Starlight deflection by the Sun, measured as $1''.7$ at the solar limb was the first evidence in favor of General Relativity. This result is often presented in contrast with so called Newtonian prediction dated back to Soldner, but in fact it announces much more by its very existence. Namely that the nature of gravity lies in geometry of spacetime shaped by distribution of mass and energy, according to the Einstein Equations. The so called Newtonian prediction is, in some sense, internally inconsistent since photons are massless so why should they feel gravity of the Sun while passing by. In relativistic context things are clear: photon's path as a geodesic in a curved spacetime is no longer a straight line.

General Relativity provided also a framework in which cosmology has been promoted from just philosophical speculations to a solid branch of physical science. Namely, we can define the goal of cosmology as studying structure of the space-time at largest scales, indeed even its global structure. First gravitational lenses were discovered in cosmological context: as multiple images of a quasar produced by foreground galaxy or as arcs around galaxy clusters representing parts of Einstein rings. Hence it is no surprise that gravitational lensing has great potential in constraining cosmological model and/or alternative theories of gravity.

This chapter starts with an introduction to the present cosmological model and its two unsolved problems of dark matter and dark energy. Especially the accelerating expansion of the Universe is a great challenge for both physics and cosmology. In light of lacking the convincing theoretical explanation, an effective description of this phenomenon in terms of a cosmic equation of state turns out useful. The strength of modern cosmology lies in consistency across independent, often unrelated pieces of evidence. Therefore, every alternative method of restricting the cosmic equation of state is important. Strongly gravitationally lensed quasar-galaxy systems create such a new opportunity by combining stellar kinematics with lensing geometry.

The utility of strongly lensed systems (galaxy and cluster lenses) for cosmology will be discussed. Then an emphasis will be put on using strong gravitational lenses as probes of cosmic equation of state which is becoming attractive in light of ongoing surveys like SLACS based on different protocols than older searches focused on potential sources. In this

approach, which has recently entered the stage of providing first estimates on cosmological parameters and will certainly develop into a technique competitive with other methods, strongly lensed systems with known central velocity dispersions act as "standard rulers" (Einstein radius being standardized by stellar kinematics).

Gravitational lenses provide also constraints on alternative theories of gravity like MOND or conformal gravity - the state of the art results in this respect will be reviewed. At last, in the present era of advances in high energy astrophysics as proclaimed by AGILE, GLAST or MAGIC projects, the discovery of lensed high energy sources appears to be a matter of not very far future. This would open new interesting opportunities. For example energy dependent time delays between images could reveal Lorentz Invariance Violation as foreseen by some approaches to quantum gravity.

Throughout this chapter the citation will many times be just indicative and by no means exhaustive, due to space limitation. Apologies go to all the authors, whose important contributions will not therefore be cited but are hereby acknowledged.

2. Standard cosmological model and its challenges

The main paradigm of modern cosmology is that geometry of the Universe can be described as one of three possible Friedman-Robertson-Walker (FRW) solutions to the Einstein equations representing homogeneous and isotropic spacetime. The FRW line element reads:

$$ds^2 = dt^2 - a(t)^2 \left[\frac{dr^2}{1 - kr^2} + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2 \right], \tag{1}$$

where, the parameter $k = -1, 0, +1$ denotes the sign of curvature of constant time hypersurfaces (identifying the aforementioned three classes of solutions – open, flat or closed). Currently there exists strong evidence, coming form independent and precise experiments, that the Universe is spatially flat. For example a combined analysis of cosmic microwave background (WMAP5), baryon acoustic oscillations (BAO) and supernova data (Hinshaw et al., 2009) gives $\Omega_{tot} = 1.0050^{+0.0060}_{-0.0061}$.

The energy-momentum tensor describing the matter content of the Universe is usually assumed in a hydrodynamical form:

$$T_{\mu\nu} = (\rho + p)u_\mu u_\nu - p g_{\mu\nu} \tag{2}$$

where: u^μ is the comoving four-velocity of the "cosmic fluid" ρ and p are the energy density and pressure, respectively.

Now, the Einstein equations get reduced to a set of two ordinary differential equations for the scale factor $a(t)$, historically known as Raychaudhuri and Friedman equation, respectively:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p), \tag{3}$$

$$H(t)^2 = \left(\frac{\dot{a}}{a} \right)^2 = \frac{8\pi G \rho}{3} - \frac{k}{a^2}. \tag{4}$$

which becomes a closed set when the equation of state $p = p(\rho)$ is specified. Usually it is assumed to be barotropic of the form $p = w\rho$. This is a very general and convenient form, containing the physically most important cases of pressureless matter ($w = 0$), radiation ($w = 1/3$) or cosmological constant ($w = -1$) but also allowing to effectively capture other

non-standard components, e.g. such as scalar fields. Energy density of the matter content of the Universe evolves according to:

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + p) = 0,$$ (5)

The Friedman equation allows for a very convenient parametrization of the expansion rate $H(t)$:

$$H(t)^2 = H_0^2[\Omega_m a(t)^{-3} + \Omega_r a(t)^{-4} + \Omega_X a(t)^{-3(1+w_X)} + \Omega_k a(t)^{-2}].$$ (6)

where Ω_i, $i \in \{m, r, X, k\}$ denote present energy density[1] of respective components (matter, radiation, other non-standard barotropic component "X" e.g. for cosmological constant we have $w_\Lambda = -1$ and Ω_Λ is just a constant term). The last term is the so called curvature term. Two Einstein equations (3),(4) describe the expansion rate and its acceleration. Their present values are known as the Hubble constant H_0 and the deceleration [2] parameter q_0.

The only gravitational degree of freedom, in the FRW cosmology, is the scale factor $a(t)$ depending on cosmic time t and responsible for temporal changes of spatial length-scales (known as cosmic expansion). Unfortunately it is not directly observable. However, there is a unique correspondence between $a(t)$ and redshift z which is an observable quantity. Namely, $a(z) = (1 + z)^{-1}$. As we mentioned above, the expansion rate $H = \frac{\dot{a}}{a}$ is determined by some set of parameters like H_0, Ω_m, Ω_r or Ω_X (if other components "X" are considered) and the equation of state parameter w_X. We will use a shorthand notation of **p** for such parameters. Technically speaking, testing cosmological models means to determine parameters **p** from observable quantities measured on samples of extragalactic objects lying far enough to feel the large-scale geometry of space-time. This specific goal of cosmology is currently called cosmography.

It is quite obvious that one very direct approach could be to test the distance – redshift relation $D(z)$ (called the Hubble diagram when plotted) whenever there is possibility to determine distances and redshifts independently. However as a consequence of non-euclidean geometry assumed, one distinguishes three types of distances in cosmology:

(i) comoving distance:

$$r(z; \mathbf{p}) = c \int_0^z \frac{dz'}{H(z'; \mathbf{p})} = \frac{c}{H_0}\tilde{r}(z; \mathbf{p})$$ (7)

where $\tilde{r}(z)$ we denotes a reduced (dimensionless) comoving distance, i.e. a comoving distance expressed as a fraction of the Hubble horizon $d_H = c/H_0$,

(ii) angular diameter distance:

$$D_A(z; \mathbf{p}) = \frac{1}{1+z}r(z; \mathbf{p})$$ (8)

(iii) luminosity distance:

$$D_L(z; \mathbf{p}) = (1+z)r(z; \mathbf{p})$$ (9)

Angular diameter distance should be used for standard rulers — the objects whose size is known a priori. It is also used in gravitational lensing theory (because gravitational lensing deals with light deflection i.e. essentially with angles). The luminosity distance is a measure invoked while using standard candles (in cosmological context: SNIa or gamma ray bursts).

---

[1] As a fraction of critical density.

[2] Strictly speaking $q_0 = -H_0^{-2}\left(\frac{\ddot{a}}{a}\right)_0$ and is called deceleration since for known forms of matter $\rho + 3p > 0$ and expansion of the Universe should decelerate.

Both distance measures are related to each other by a $(1 + z)^2$ factor (see above) which is known as Etherington reciprocity relation.

2.1 Dark matter problem

Cosmology as it evolved, became able to pose meaningful physical questions concerning processes which occurred in the early Universe (baryogenesis, nucleosynthesis, radiation domination and recombination epochs etc.) and study the large scale structure (LSS) formation. In early days of physical cosmology (in the 70-ties of the XX century) observational capabilities of those days defined the goal of cosmology as the "quest for two numbers H_0 and q_0" (Sandage, 1970). In the present era of precise observations of cosmic microwave background (CMB) anisotropies (starting with COBE, BOOMERANG, MAXIMA, to WMAP and the present PLANCK mission), LSS studies from deep surveys (like 2dFRGS, SDSS, 2MASS) the improved light elements (^{2}D, ^{3}He, ^{4}He, ^{7}Li) abundances assessments, when the connection between cosmology and particle physics became more intimate than ever before, our knowledge of physics and evolution of the Universe is captured in more than 20 parameters determined with high precision e.g. from WMAP data (Spergel et al., 2003).

One of, then and now important, such parameters is the matter density Ω_m. This is the point at which the problem of dark matter comes into play in a very consistent way. Namely, all present matter density determinations coherently point toward $\Omega_m = 0.25 - 0.30$. However, probes such like CMB anisotropies (Dunkley et al., 2009), baryon acoustic oscillations (BAO) (Percival et al., 2007) or supernovae Ia provide fits on Ω_m entangled with H_0. The only method sensitive exclusively to matter density comes form studying peculiar velocities of galaxies. The analysis of Feldman et al. (2003) gave $\Omega_m = 0.30^{+0.17}_{-0.07}$. Later Mohayaee & Tully (2005) applied orbit retracing methods to motions in the local supercluster and obtained $\Omega_m = 0.22 \pm 0.02$, which is also consistent with the range reported above. On the other hand, primordial nucleosynthesis restricts the baryonic matter content to $\Omega_b = 0.04 \pm 0.01$ assuming $h = 0.7$ (Cyburt, 2004). This means that most of gravitationally interacting matter at scales of galaxies and clusters should be of non-baryonic nature and is commonly known under the name of the dark matter.

The case for dark matter was strengthened in several independent ways. First of all, and this was historically the first time when dark matter was invoked, dynamical studies of spiral galaxies revealed flat rotation curves extending far beyond the region from which most of the galactic light is emitted. Contrary to the expectations it showed that galaxies are about 10 times more massive than their luminous components. Hence the dark matter problem is also called the missing mass problem. Elliptical galaxies do not have such good rotation tracers as the spirals, but velocity dispersions of theirs stars support the claim that they are more massive than luminous components. Then gravitational lensing provides independent way of proving that ellipticals contain dark matter. X-ray observations of galaxy clusters revealed the existence of hot intergalactic gas whose temperature reflects the gravitational potential by virtue of hydrostatic equilibrium. On one hand it aslo confirmed the reality of the "missing mass", independently supported by cluster mass determination from cluster strong lensing manifested as giant arc around the clusters.

The case of the so called bullet cluster - a structure in which smaller sub cluster undergone a collision with a larger galaxy cluster 1E 0657-56 is often reported as a smoking gun evidence for dark matter (Clowe et al., 2006). Namely the mass profiles reconstructed by weak lensing techniques (sensitive to the total mass distribution) and by X-ray data reveal a characteristic mismatch between the total mass (anchored in the main cluster and a sub-cluster, respectively)

and the intergalactic gas – stripped from clusters during the collision and left compressed in between. Similar behavior i.e. that dark matter and stellar systems passed through the collision while the intergalactic gas interacted and was stripped, has been found in the cluster MACS J0025.4-1222 (Bradač, 2008).

The notion of dark matter is independently supported, form theoretical pint of view, by particle physicists. They have for a long time contemplated the extensions of the standard model of particle physics and provided a long list of potential candidates for dark matter particles, like supersymmetric particles (e.g. neutralino), axion, Kaluza-Klein gravitons etc. All this evidence is so compelling that many experiments aimed at direct detection of dark matter particles are underway. However, there exist several facts creating problems to this picture. For example the large scale structure structure formation simulations typically predict much more structure at small scales than is actually seen. Observational fact that large galaxies are typically surrounded by their dwarf companions qualitatively fits this picture, but their abundances are by far too low with respect to predictions. The solution might be that there are indeed clumps of dark matter around galaxies, which had not anchored enough gas for star formation. Because gravitational lensing is sensitive to mass irrespective of its nature (luminous - bayonic or dark), there is much hope that gravitational lensing studies will help to settle the issue.

2.2 Dark energy problem

One of the most important issues in modern cosmology – in fact having consequences for the fundamental physics – is the problem of currently accelerating expansion of the Universe, known as the dark energy problem. It was first convincingly demonstrated in Hubble diagrams obtained from the SN Ia data e.g. (Amanullah et al., 2010; Perlmutter et al., 1999; Riess et al., 1998) and later supported by other independent studies including: CMBR anisotropies (Hinshaw et al., 2009; Komatsu et al., 2009) or baryon acoustic oscillations imprinted in the large scale structure power spectrum (Eisenstein & Hu, 1999; Percival et al., 2007). This result was surprising, since as can be seen from the Einstein equations (3) and (4) all known forms of matter (having the trace of energy-momentum tensor positive) should decelerate the expansion.

The explanation of this phenomenon is far from obvious and presents a fundamental challenge to standard models of both particle physics and cosmology. The new physics of dark energy may lie in the nature of gravity, the quantum vacuum, or extra dimensions. Concerning the first possibility there exists an increasing body of literature, e.g. (Buchert, 2001; Wiltshire, 2007) pointing out that if one attempts to average out local sources of gravity (galaxies and clusters) in order to obtain the smooth description of the Universe, such averaging procedure, not commuting with temporal evolution, could manifest as an additional source term in the energy-momentum tensor. Within the second possibility our ideas about the quantum vacuum are expressed by either introducing cosmological constant Λ or some scalar field evolving in time (quintessence). The last possibility is to contemplate modifications to the Friedman-Robertson-Walker models arising e.g. in brane-world scenarios.

2.3 Different scenarios of dark energy

The ΛCDM model is a FRW cosmology with non-vanishing cosmological constant and pressure-less matter including the dark matter component. It is a standard reference point in modern cosmology also called the concordance model since it fits rather well to independent

data (such like CMBR data, Large Scale Structure considerations, supernovae data). There are, however, reasons why we are not fully satisfied with the concordance scenario. First, the cosmological constant suffers from the fine tuning problem: being constant, why does it start dominating at the present epoch? Next, if we imagine its origin as the quantum-mechanical energy of the vacuum, field theoretical estimates predict its value 120 orders of magnitude larger than observed — the biggest discrepancy of theoretical physics today (Weinberg, 1989). Therefore the next, popular explanation of the accelerating Universe is to assume the existence of a negative pressure component called dark energy. One can heuristically assume that this component is described by hydrodynamical energy-momentum tensor with (effective) cosmic equation of state: $p = w\rho$ where $w < -1/3$. In such case this component is called "quintessence". Usually the quintessence is attributed to some sort of a scalar field. Another scalar field invoked by cosmologists is the inflaton, which in order to accomplish its role as driving the inflation and creating particles at the reheating epoch, clearly had its own dynamics. Therefore thinking about quintessence as having origins in the evolving scalar field, would lead to a natural expectation that w coefficient should vary in time, i.e. $w = w(z)$. Bearing in mind that the scale factor $a(t)$ is a real physical degree of freedom instead of the redshift z, the parametrization of $w(z) = w_0 + w_a \frac{z}{1+z}$ developed by Chevalier & Polarski (2001) and Linder (2003) turned out to be well suited for such case. For the purpose of providing an effective description of dark energy testable with real data it is sufficient to stop here, since the numerous particular scenarios (see e.g. Caldwell & Kamionkowski (2009)) – stemming from different physical inspirations – all have common point of empirical confrontation in the w coefficient and its temporal evolution. Two more specific models deserve however a mention: the Chaplygin gas and brane-world model of Dvali, Gabadadze and Porrati. The formulae for expansion rates in these models are given in Table 1.

Model	Cosmological expansion rate $H(z)$ (the Hubble function)
ΛCDM	$H^2(z) = H_0^2 \left[\Omega_m (1+z)^3 + \Omega_\Lambda \right]$
Quintessence	$H^2(z) = H_0^2 \left[\Omega_m (1+z)^3 + \Omega_Q (1+z)^{3(1+w)} \right]$
Chevalier-Polarski-Linder	$H^2(z) = H_0^2 \left[\Omega_m (1+z)^3 + \Omega_Q (1+z)^{3(1+w_0+w_a)} \exp(\frac{-3w_a z}{1+z}) \right]$
Chaplygin Gas	$H^2(z) = H_0^2 \left[\Omega_m(1+z)^3 + \Omega_{Ch} \left(A_0 + (1-A_0)(1+z)^{3(1+\alpha)} \right)^{\frac{1}{1+\alpha}} \right]$
Braneworld	$H^2(z) = H_0^2 \left[(\sqrt{\Omega_m(1+z)^3 + \Omega_{r_c}} + \sqrt{\Omega_{r_c}})^2 \right]$

Table 1. Expansion rates $H(z)$ in cosmological models representative to various dark energy scenarios.

In the class of generalized Chaplygin gas models matter content of the Universe consists of pressure-less gas with energy density ρ_m representing baryonic plus cold dark matter (CDM) and of the generalized Chaplygin gas with the equation of state $p_{Ch} = -\frac{A}{\rho_{Ch}{}^\alpha}$ representing dark energy responsible for acceleration of the Universe. The original Chaplygin gas corresponds to $\alpha = 1$. In cosmological context it has been promoted to the role of free parameter in phenomenological approach. Values of α exponent close to zero mean that the model is equivalent to ΛCDM case. This exotic form of cosmic equation of state is inspired by some super-string theories, but at the phenomenological level it has an advantage that it smoothly interpolates the expansion history of the Universe from matter dominated to dark energy dominated regimes. Chaplygin models have been confronted with different cosmological data like supernovae (Biesiada et al., 2005), cosmic microwave background radiation anisotropies (Amendola et al., 2003) or baryonic acoustic oscillations (Wu & Yu, 20).

At last, the brane-world models belong to the class of theories which seek the solution of presently accelerating expansion of the Universe not in an exotic material component, but in modifications of gravity. Brane-world scenarios assume that our four-dimensional spacetime is embedded into 5-dimensional space and gravity in 5-dimensions is governed by the usual 5-dimensional Einstein-Hilbert action. The bulk metric induces a 4-dimensional metric on the brane. The brane induced gravity models (Dvali et al., 2000) have a 4-dimensional Einstein-Hilbert action on the brane calculated with induced metric. According to this picture, our 4-dimensional Universe is a surface (a brane) embedded into a higher dimensional bulk space-time in which gravity propagates. Therefore there exists a certain cross-over scale r_c above which an observer will detect higher dimensional effects.

As a consequence of modified gravity, the Friedman equation reads

$$H^2 + \frac{k}{a^2} = \left[\sqrt{\frac{\rho}{3M_{Pl}^2} + \frac{1}{4r_c^2}} + \frac{1}{2r_c} \right]^2$$

from which the expansion rate function shown in Table 1 can be derived in flat (i.e. $k = 0$.) case. In flat brane-world Universe the following relation is also valid: $\Omega_{r_c} = \frac{1}{4}(1 - \Omega_m)^2$. Cosmological models in brane-world scenarios have been widely discussed in the literature. Quite recent paper (Xu & Wang, 2010) presents one of the most comprehensive analysis of brane-world models by considering jointly the data from supernovae, gamma-ray bursts, BAO, CMB peaks, the look back times and growth functions for the large scale structure. Their results (posterior probability distributions for model parameters) obtained by using the Markov Chain Monte Carlo simulation yield $\Omega_m = 0.266^{+0.0298}_{-0.0304}$

The potential of constraining dark energy models with SNIa data, even though ever increasing, would not be sufficient if taken alone in separation form the other approaches. Indeed, the power of modern cosmology lies in building up consistency rather than in single, precise, crucial experiments. Therefore, every alternative method of restricting cosmological parameters is desired. In this spirit a number of combined analyses involving CMBR measurements, age-redshift relation, x-ray luminosities of galaxy clusters or the large scale structure considerations have been performed in the literature. It is exactly in this context that cosmography performed on strong lensing systems is becoming useful.

3. Gravitational lensing in brief

Gravitational lensing of quasars and extragalactic radio sources at high redshifts by foreground galaxies is now well established and has developed into a mature branch of both theoretical and observational astrophysics. For comprehensive review and introduction to the theory of gravitational lensing see e.g. Schneider et al. (1992) or Schneider et al. (2006).

Imagine the source, observer and some other massive object (later to be called lens) located exactly along a line. From the point of view of traditional optics the source would be obscured by the intervening object: the only light ray (or a small collimated bunch of rays) pointing toward the observer would not reach him. General relativistic phenomenon of light deflection near massive bodies changes this picture: out of all light rays emitted radially some (going close to the deflector – how close it depends on the mutual locations of source, deflector and observer) are now focused at the observer. Without General Relativity they would have missed him. Intervening massive body acts as a lens and a source behind reveals its existence as a luminous ring — the so called Einstein ring. Even the smallest misalignment of the

source, the lens and observer results typically in multiple images whose angular positions and magnification ratios allow reconstructing lensing mass distribution.

As in ordinary optics, there are two equivalent approaches to understand the phenomenon: the light rays formalism and the wavefronts formalism. From the point of view of Fermat's principle, the light travel time can be calculated as

$$t(\mathbf{x}) = \frac{1 + z_l}{c} \frac{D_l D_s}{D_{ls}} \left[\frac{1}{2} (\mathbf{x} - \beta)^2 - \psi(\mathbf{x}) \right] \tag{10}$$

where: $\mathbf{x}$ and β are positions (as projected on the celestial sphere) of the image and the source, D_l, D_s are angular diameter distances to the lens and the source located at redshifts z_l and z_s respectively (D_{ls} is the angular diameter distance between lens and source). $\psi(\mathbf{x})$ is the projected gravitational potential (i.e. the actual potential integrated along line of sight) satisfying two dimensional Poisson equation:

$$\triangle \psi = 2\kappa \tag{11}$$

where κ is the (projected) surface mass density in units of critical density $\Sigma_{cr} = c^2 D_s / (4\pi G D_l D_{ls})$. Then, the Fermat's principle states that images form at stationary points of time delay surface $\nabla t(\mathbf{x})$, which leads to the lens equation:

$$\beta = \mathbf{x} - \nabla \psi = \mathbf{x} - \alpha \tag{12}$$

The last equality is usually invoked in the light-rays formalism, where $\alpha = \frac{D_{ls}}{D_s} \hat{\alpha}$ is scaled deflection angle. In axially symmetric lenses, for example: $\hat{\alpha}(\mathbf{x}) = \frac{4GM(x)}{c^2 x^2} \mathbf{x}$ where $M(x)$ is the mass enclosed by the circle of radius $x = |\mathbf{x}|$. The most useful notion in gravitational lensing theory is the Einstein radius ϑ_E. In circular lenses it is the radius of the circle inside which the average projected mass density is equal to critical density (cf. above). For non circular lenses this should be modified appropriately. Thus the Einstein radius defines the deflection scale of a given lens.

The lensing is called strong if source position happens to lie within the circle of a radius ϑ_E. In this case multiple images appear. In the opposite case (i.e. the light-rays from the source passing by the lens outside its Einstein radius) there are no multiple images. However even in this case light-ray bundle experiences systematic distortion which changes the shape of the lensed source. This phenomenon is called weak lensing, has its own place in cosmology (Schneider et al., 2006) and is beyond the scope of this chapter.

Since lensing galaxies are often ellipticals, the number of images is usually equal to four – the issue of image multiplicity is discussed e.g. in (Schneider et al., 1992). However, the surprisingly realistic model of the lens potential is that of a singular isothermal sphere (SIS) in which the 3-dimensional mass density has the following profile:

$$\rho = \frac{\sigma_{SIS}^2}{2\pi G r^2} \tag{13}$$

Indeed lensing by ellipticals can be modeled by its variant called singular isothermal ellipsoid (SIE). Therefore for the illustrative purposes it would be sufficient to restrict our attention to the SIS model. Other realistic and more sophisticated models are discussed in classical textbooks (Schneider et al., 1992).

The Einstein ring radius for the SIS model is:

$$\vartheta_E = 4\pi \frac{D_{ls}}{D_s} \frac{\sigma^2}{c^2} \tag{14}$$

where σ denotes one-dimensional velocity dispersion of stars in lensing galaxy. If the lensing is strong i.e. $\beta < \vartheta_E$ then two co-linear images A and B form on the opposite side of the lens, at radial distances $R_A = \vartheta_E + \beta$ and $R_B = \vartheta_E - \beta$.

Besides multiple images, another important ingredient of gravitational lensing is the time delay between lensed images of the source. This effect originates as a competition between Shapiro time delay from the gravitational field and the geometric delay due to bending the light rays and is best understood in terms of Fermat principle. In other words, the intervening mass between the source and the observer introduces an effective index of refraction, thereby increasing the light travel time. In the aforementioned SIS model, time delay between the images is:

$$\Delta t_{SIS} = \frac{1 + z_l}{2c} \frac{D_l D_s}{D_{ls}} (R_A^2 - R_B^2) \tag{15}$$

which according to the above mentioned relations for SIS model can also be written as

$$\Delta t_{SIS} = \frac{2(1 + z_l)}{c} \frac{D_l D_s}{D_{ls}} \vartheta_E \beta = \frac{8\pi}{H_0} \tilde{r}_l \beta \frac{\sigma^2}{c^2} \tag{16}$$

In the last equation $\tilde{r}_l$ denotes the reduced comoving distance to the lens. The equation (15) is commonly used by gravitational lensing community because it reduces time delay problem to relative astrometry of images, whereas β is much harder to asses (it is small in order for strong lensing to occur) and Einstein ring radius is not a directly observable quantity (although image separation fairly represents the Einstein radius). However, the equation (16) is more useful from the theoretical point of view. In particular it shows explicitly that the time delay between images is created at the lens location ($\tilde{r}_l$ factor).

The last observable derivable from strongly lensed systems is the flux ratio of images. It is the most sensitive with respect to details of mass distribution along the light-ray path (both in terms of detailed knowledge of smooth component of mass distribution as well as the graininess of the lens i.e. microlensing by stars or other clumped massive structures along the path.)

Let us close this brief introduction with the reminder that gravitational lensing, as general relativistic effect, is achromatic. It means that the locations of images and the time delay do not depend on the wavelength of light in which they are observed.

4. Cosmography with strong lensing systems.

4.1 Brief history

Cosmographic application of strong lensing systems has quite a long history. Since the discovery of the first gravitational lens the number of strongly lensed systems increased to a hundred (in the CASTLES database [3]) and is steadily increasing following the new surveys like SLACS (Sloan Lens ACS Survey [4]) and the opportunity for gravitational lenses being competitive to other techniques becomes real.

---

[3] http://www.cfa.harvard.edu/castles/
[4] http://www.slacs.org/

The first theoretical proposal of serious application of strong gravitational lensing was presented by Refsdal (1964) in his stimulating paper on the Hubble constant measurements from time delays between images. Namely if the lensed source is intrinsically variable (quasars being the main population of sources display such variability) and we are able to extract the variability pattern from the light-curve (which in practice is non-trivial task), this variability would be observed at different times in the images. Then the time-delay, as e.g. (15) depends on image locations and relative distances of the source, lens and observer. But the magnitude of this delay (temporal scale of the effect) is given by H_0^{-1}. This creates alternative possibility of measuring the Hubble constant H_0 which is, unlike the other methods, independent of the cosmic distance ladder and its calibration. The number of lenses with reliably measured time delays has accumulated slowly over decades. Several years ago there were about 10 such lenses and the observational status of the Hubble constant determination, as reviewed in details in (Schneider et al., 2006), was that time delays preferred $H_0 = 52 \pm 6 \ km \ s^{-1} \ Mpc^{-1}$ in contrast to the HST Key Project value of $H_0 = 72 \pm 8 \ km \ s^{-1} \ Mpc^{-1}$ (Freedman et al., 2001). Later papers of Oguri (2007) and Coles (2008) announced the results ($H_0 = 68 \pm 16 \ km \ s^{-1} \ Mpc^{-1}$ and $H_0 = 71 \pm 8 \ km \ s^{-1} \ Mpc^{-1}$ respectively) now in agreement with HST value.

Other cosmological parameters, like Ω_m, Ω_Λ or w coefficient in quintessence models up to now have been assessed on samples drawn from radio survey Cosmic Lens All Sky Survey (CLASS)– Chae (2007) and SDSS Quasar Lens Search (SQLS) – Oguri et al. (2008). The approach they taken was to assess the frequency of lensed systems given a source population. This quantity depends on the cosmic volume over which sources are distributed, which makes it useful for cosmography. However, there are other factors influencing the result such as reliable lens model adopted, contribution from the line of sight contamination by intervening massive structures or source luminosity function. In the next section we will discuss another technique of using strong lenses as tools for cosmography.

4.2 Testing dark energy models: strong lensing systems as standard rulers

The first formulations of this approach can be traced back to Futamase & Yoshida (2001). Then, the idea of using strong lensing for measuring the cosmic equation of state was discussed in Biesiada (2006) and also in a later paper by Grillo et al. (2008).

The main idea is that formula for the Einstein radius in a SIS lens

$$\theta_E = 4\pi \frac{\sigma_{SIS}^2}{c^2} \frac{D_{ls}}{D_s} \tag{17}$$

depends on the cosmological model through the ratio of (angular-diameter) distances between lens and source and between observer and lens. The angular diameter distance in flat Friedmann-Robertson-Walker cosmology is given by (8).

Provided one has reliable knowledge about the lensing system: i.e. the Einstein radius θ_E (from image astrometry) and stellar velocity dispersion σ_{SIS} (form central velocity dispersion σ_0 obtained from spectroscopy) one can use it to test the background cosmology. This method is independent on the Hubble constant value (which gets canceled in the distance ratio) and is not affected by dust absorption or source evolutionary effects. It depends, however, on the reliability of lens modeling (e.g. SIS or SIE assumption) and measurements of σ_0. Hopefully, starting with the Lens Structure and Dynamics (LSD) survey and the more recent SLACS survey spectroscopic data for central parts of lens galaxies became available allowing to assess their central velocity dispersions. There is a growing evidence for homologous structure

of late type galaxies (Koopmans et al., 2006; 2009; Treu et al., 2006) supporting reliability of SIS/SIE assumption. In particular it was shown there that inside one effective radius massive elliptical galaxies are kinematically indistinguishable from an isothermal ellipsoid.

In the method outlined above cosmological model enters not through a distance measure directly, but rather through a distance ratio

$$\mathcal{D}^{th}(z_l, z_s; \mathbf{p}) = \frac{D_s(\mathbf{p})}{D_{ls}(\mathbf{p})} = \frac{\int_0^{z_s} \frac{dz'}{h(z'; \mathbf{p})}}{\int_{z_l}^{z_s} \frac{dz'}{h(z'; \mathbf{p})}} \tag{18}$$

and respective observable counterpart reads:

$$\mathcal{D}^{obs} = \frac{4\pi\sigma_0^2}{c^2\theta_E}$$

This has certain consequences both advantageous and disadvantageous. The positive side is that the Hubble constant H_0 gets canceled, hence it does not introduce any uncertainty to the results. On the other hand we have a disadvantage that the power of estimating Ω_m is poor (which could be seen by inspection into specific formulae for $h(z; \mathbf{p})$ – see Table 1). Cosmological model parameters (coefficients in the equation of state, in particular) can be estimated by minimizing the chi-square:

$$\chi^2(\mathbf{p}) = \sum_i \frac{(\mathcal{D}_i^{obs} - \mathcal{D}_i^{th}(\mathbf{p}))^2}{\sigma_{\mathcal{D},i}^2} \tag{19}$$

where the sum is over the sample and $\sigma_{\mathcal{D},i}^2$ denotes the variance of $\mathcal{D}^{obs}$ (contextual use of the same symbol for variances and velocity dispersions should not lead to confusion). Putting aside the issue that the observable quantity here is a distance ratio, one can see that strong lenses constitute a class of standard[5] rulers.

The above method extensively investigated by Grillo et al. (2008) on simulated data was first used in practice to constrain various cosmological models in Biesiada et al. (2010) where ΛCDM, quinetssence and CPL model were constrained. Later it was used (together with SNIa, CMB and BAO data) as a part of joint analysis in (Biesiada et al., 2011). Table 2 summarizes recent constraints on parameters of the cosmological model obtained with strong gravitational lensing.

The results obtained were generally in agreement with those obtained by other authors with different methods. In particular at the 2σ level they agree with the supernovae Ia results. One should note however that in the CPL model (quintessence with time varying equation of state) standard rulers (strong lenses in particular) display a systematic shift downwards in (w_0, w_a) plane with respect to supernovae Ia. Such shift in best fitted parameters inferred from supernovae (standard candles, sensitive to luminosity distance) and BAO or acoustic peaks (standard rulers, sensitive to angular diameter distance) has already been noticed and discussed by Lazkoz et al. (2009) and by Linder & Roberts (2008). Bearing in mind similar mutual inconsistency in the Hubble constant values inferred from lensing time delays and

---

[5] They could be better called "standardizable" rulers because each lens has intrinsically different Einstein radius, but stellar kinematics, i.e. velocity dispersion allows for disentangling the effect of mass from that of distances. In a sense supernovae Ia could also be called "standardizable" not standard candles, since it is the stretch factor correction that makes them standard.

Cosmological model	Strong lenses	Lenses + BAO + R	SN Ia
ΛCDM	$\Omega_m = 0.266 \pm 0.280$	$\Omega_m = 0.273 \pm 0.018$	$\Omega_m = 0.275 \pm 0.020$
Quintessence	$\Omega_m = 0.27$ prior	$\Omega_m = 0.262 \pm 0.035$	$\Omega_m = 0.299 \pm 0.075$
	$w = -0.632 \pm 0.446$	$w = -1.066 \pm 0.188$	$w = -1.070 \pm 0.215$
Chevalier-Polarski-Linder	$\Omega_m = 0.27$ prior	$\Omega_m = 0.276 \pm 0.055$	$\Omega_m = 0.228 \pm 0.156$
	$w_0 = 0.359 \pm 1.245$	$w_0 = -0.824 \pm 0.704$	$w_0 = -0.993 \pm 0.207$
	$w_a = -3.630 \pm 5.328$	$w_a = -0.757 \pm 2.148$	$w_a = 0.609 \pm 1.071$
Chaplygin Gas		$\Omega_m = 0.273 \pm 0.018$	$\Omega_m = 0.275 \pm 0.020$
		$A = 1.000 \pm 0.001$	$A = 0.999 \pm 0.004$
		$\alpha = -0.040 \pm 2.260$	$\alpha = 0.006 \pm 0.372$
Braneworld		$\Omega_m = 0.345 \pm 0.021$	$\Omega_m = 0.117 \pm 0.015$

Table 2. Constraints on the parameters of various cosmological model obtained with strong lensing systems. First column refers to a sample of strong lenses only (Biesiada et al., 2010), the middle refers to lenses analysed jointly with other standard rulers, whereas the last column shows fits with Union2 SNIa data (after Biesiada et al. (2011))

from the HST Key Project (Schneider et al., 2006), this result suggests the need for taking a closer look at compatibility of results derived by using angular diameter distances and luminosity distances respectively. The ideas of testing the Etherington reciprocity relation between these two distance measures have been discussed by Basset & Kuntz (2004) and by Uzan et al. (2005).

Although the sample of suitable lenses (i.e. with good measurements of Einsetin radii, source and lens redshifts and central velocity dispersions) has been rather small ($n = 20$ lenses) the ongoing SLACS survey is providing new strong lensing systems which is very encouraging for further applications of the method. The strategy adopted in SLACS survey is particularly important. The earlier searches were focused on source population (quasars) seeking for close pairs or multiples and checking if they are multiple images of a single source lensed by an intervening galaxy. Therefore a high lensing probability was an important selection factor there.

Since lensing probability is proportional to the area of the Einstein ring, it means that two factors are crucial in this context. First, is the mass of the lens. This is the main reason why in vast majority of cases the lens is E/S0 type galaxy. This could be understood since ellipticals being a latecomers in hierarchical structure formation are created in mergers of low-mass spiral galaxies. Hence they are more massive than spirals and the probability of their acting as lenses is higher. Second factor is the distance ratio D_{ls}/D_s. In details, this of course depends on the cosmological model, but it is maximal when the lens is located roughly half way between the source and the observer. The SLACS sample has an average D_{ls}/D_s ratio equal to 0.58 with an rms scatter 0.15 (Treu et al., 2006). Whereas for their purpose (investigating galactic dynamics with strong lenses) it was advantageous, in our context it weakens the performance of the method. Therefore having a sub-sample of lenses with the distance $\mathcal{D}$ ratio deviating from the mean more than rms in either direction would be beneficial and in this respect SLACS survey is encouraging. Namely, the SLACS survey is focused on possible lens population (massive ellipticals) with good spectroscopic data. Using SDSS templates spectra are carefully checked for residual emission (at least three distinct common atomic transitions) coming from higher redshifts. Such candidates undergo image processing by subtracting parametrized brightness distribution typical for early type galaxies in order to reveal multiple images of the quasar. Details can be found in Bolton et al. (2006). Therefore,

besides the obvious bonus of having central velocity dispersion measured, such strategy is better suited for discovering systems with larger D_{ls}/D_s ratios which in turn can be used for testing cosmological models.

4.3 Cosmography with cluster lensing

Besides the galaxies acting as lenses, their clusters – first virialized structures in the Universe – do the same. The cores of galaxy clusters have surface densities which are typically much larger than the critical surface density Σ_{cr} for multiple image production. Therefore they are able to produce strongly lensed images of galaxies and quasars lying behind them. Such images manifest themselves as luminous arcs around clusters. Historically it was Paczyński (1987) who proposed that giant arcs might be gravitationally lensed images of background galaxies. First measurements of arcs' redshifts proved this definitely. The possibility of constraining cosmology with CSL systems has been explored in the past e.g. (Paczyński & Górski, 1981; Sereno, 2002) and still remains a fruitful, fast developing field of research. It is typical that we observe multiple sets of arcs in cluster lenses corresponding to different sources (with different redshifts) lensed by the same cluster. Hence, the abundance of arcs may provide useful cosmological constraints, in a manner similar to the statistics of multiple images in galaxy lenses. For example, Meneghetti et al. (2005) explored the statistics of arcs in various cosmological models and found that arc abundances can be used to differentiate between dark energy scenarios. More recently, Gilmore & Natarayan (2009) explored the prospect of combining cluster lensing systems as a more powerful probe of dark energy.

Analogously to the method outlined in the preceding section, the locations of images in cluster lensing systems also contain useful cosmological information. Namely, the image positions depend not only on the mass distribution, but also on the angular diameter distances between the observer, lens, and source. If more than one set of images is observed, the geometrical dependence may be exploited to probe the cosmological parameters even with a single cluster lens. One of the best studied cluster lensing system is Abell 1689. The mean redshift of this cluster is $z_l = 0.184$ and it is one of the richest clusters in terms of the number density of galaxies in its core. In a recent paper by Jullo et al. (2010) this cluster was used to derive constraints on the cosmological parameters Ω_m and w. Based on images from the Advanced Camera for Surveys (ACS) this cluster is known to produce 114 multiple images from 34 unique background galaxies. This allowed Jullo et al. (2010) to use many observables like (18) from a single cluster. To be more specific instead using $\mathcal{D}^{th}$ like in (18), they used quotients formed pairwise for background sources

$$\mathcal{D}_{cl}^{th} = \frac{\mathcal{D}^{th}(z_l, z_{s1}; \mathbf{p})}{\mathcal{D}^{th}(z_l, z_{s2}; \mathbf{p})} \tag{20}$$

where: z_l is the cluster's redshift, z_{s1} and z_{s2} are reshifts of respective pair of sources.

Applying the following criteria: demand of good spectroscopic data for images and excluding regions where mass reconstruction gets poorer, from the initial 114 images, Jullo et al. (2010) selected finally 28 images which they further used to constrain cosmological parameters to $\Omega_m = 0.25 \pm 0.05$ $w = -0.97 \pm 0.07$.

Even more promising is the idea of using a larger sample of cluster lensing systems. Such an approach has the advantage that results obtained from different lines of sight are statistically independent. As discussed by Gilmore & Natarayan (2009) competitive constraints can be obtained by combining at least 10 lenses with 5 or more image systems. We may therefore conclude that cluster strong lensing is becoming a very useful complementary tool,

particularly in probing dynamic dark energy models, which demand larger range of redshifts to be probed.

5. Constraining alternative theories of gravity with strong lensing

Despite the great successes of General Relativity, and Newtonian gravity (if taken in appropriate limits), one should be aware that the majority of its direct tests have been within the Solar system. Measurements of time delays in binary pulsar systems (Taylor et al., 1992) (Taylor et al. 1992) have also verified General Relativity, but all such tests were performed in large acceleration regimes (large as compared to accelerations e.g. experienced by stars at the outskirts of galaxies). Therefore, the possibility remains that gravity at large distances (or small accelerations) could be different. Similar reasoning aimed at explaining dark energy, underlaid, an already discussed, idea of brane-world cosmologies. In the context of dark matter (or "missing mass") problem the most famous idea of this kind was formulated by Milgrom (1983) and is known as MOND. MOND is able to explain (or at least not to be in conflict with) all kinematical properties of stellar systems only in terms of their baryonic constituents (stars and gas). More specifically, Milgrom introduced a preferred scale of acceleration $a_0 \approx 10^{-10} \ ms^{-2}$ of the order of the centripetal accelerations of stars and gas clouds in the outskirts of disk galaxies, and postulated a modified form of the second law of dynamics: $m\mu(a/a_0)\mathbf{a} = \mathbf{F}$ where: $\mu(x)$ is certain function (not specified) which has appropriate asymptotic form: $\mu(x) = x$ for $x << 1$ and $\mu(x) = 1$ for $x >> 1$ in order to interpolate between high acceleration (newtonian) regime and low acceleration (MONDian) regime. In terms of it MOND relates the acceleration $\mathbf{a}$ of a test particle to the Newtonian gravitational field $-\nabla\Phi_N$ generated by the baryonic mass density alone by

$$\mu(a/a_0)\mathbf{a} = -\nabla\Phi_N$$

Designed to solve the flat rotation curve, it also predicted a lot of phenomenological relations such like Tully-Fisher relation and is successful in explaining the shapes of rotation curves. It is amazing that single value of $a_0 = 1.2 \ 10^{-10} \ m/s^2$ reproduces both rotation curves and Tully-Fisher relation in over a hundred galaxies. Excellent reviews of other phenomenological successes of MOND can be found in (Milgrom, 2008; Sanders & McGaugh, 2002).

On the other hand, it is observational fact that in all cases we know, the "missing mass" discovered kinematically manifests itself also in enhanced gravitational lensing power. It means that gravitational lensing could be used to test alternative theories like MOND. However, gravitational lensing is a relativistic effect, while MOND in its original form is not. Therefore there was a lot of effort devoted to construct proper relativistic extension of MOND, which culminated in formulation of TeVeS theory by Bekenstein (2004). TeVeS formulates the gravity theory in terms of three fundamental fields: tensor (the metric), vector and scalar, later it was shown that it could be reduced to just tensor-vector theory – for the state of art review on this see Bekenstein (2010).

One of first considerations of gravitational lensing in MOND were due to Mortlock & Turner (2001). They rightly identified two limiting regimes of lensing: high acceleration (so called Newtonian) where the deflection angle agrees with relativistic predictions and low acceleration (deep MOND) regime where the deflection angle approaches the constant value. Deep MOND regime is valid for light-ray impact parameters $b >> b_0 := (k/2\pi)(GM/a_0)^{1/2}$. This result has been confirmed by Chiu et al. (2006), this time consequently in the TeVeS framework. However, they also demonstrated that in the intermediate MOND regime $b \approx b_0$ the deflection angle depends crucially on the details of the theory (e.g. specification of $\mu(x)$

function in MOND). Besides they demonstrated that in point lens model, the difference in image amplifications is no longer unity, as in General Relativity, and that it depends on the mass of the lens. They also investigated the gravitational time delay in TeVeS which is important for interpreting differential time delays in doubly imaged variable quasars. The baryon distribution in a galaxies is no longer point-like but can be better represented by e.g. the Hernquist profile. Zhao et al. (2006) employed it to compare TeVeS predictions with a large sample of quasars doubly imaged by intervening galaxies and obtained encouraging (in favor of TeVeS) results. The weak lensing by clusters of galaxies, presents a problem for pure MOND. Especially the bullet cluster (a pictorial case for dark matter) was claimed to kill the MOND approach. After first encouraging approaches to explain it in TeVeS, it was established that it cannot be done with purely baryonic matter distribution – a collisionless component of massive 2 eV neutrinos is necessary (Angus et al., 2006).

6. Strong lensing tests of Lorentz Invariance Violation

Despite the fact that quantum gravity theory still remains elusive, it is generally expected that it will bring the picture of a space-time foam at short distances leading to Lorentz Invariance Violation (LIV) manifested e.g. by energy dependent modification of standard relativistic dispersion relation (Amelino-Camelia et al., 1998; Amelino-Camelia & Piran, 2001).

Several years ago it has been proposed to use astrophysical objects to look for energy dependent time of arrival delays. Specifically gamma ray bursts (GRBs) being highly energetic events visible from cosmological distances are the most promising sources of constraining LIV theories (Ellis et al., 2006). Among other sources the BL Lac objects like Mk 501 are considered. It is this particular object from which 20 TeV photons were reported (Amelino-Camelia & Piran, 2001). Such objects (also called blazars) have similar nature with quasars. The idea of searching for time of flight delays is tempered however by our ignorance of intrinsic delay (at source frame) in different energy channels. In (Biesiada & Piórkowska, 2009) a test based on gravitational lensing, which free from this limitation has been proposed. In the rest of this section we will follow it closely.

Let us consider a phenomenological approach to LIV theories by assuming the modified dispersion relation for photons in the form (Amelino-Camelia & Piran, 2001):

$$E^2 - p^2c^2 = \epsilon E^2 \left(\frac{E}{\xi_n E_{QG}} \right)^n \tag{21}$$

where: $\epsilon = \pm 1$ is the so called "sign parameter", ξ_n is a dimensionless parameter associated with energy scale at which n-th order corrections to the dispersion relation become important. As a first guess (having no other suggestions) one may assume E_{QG} equal to the Planck energy, then: $\xi_1 = 1$ and $\xi_2 = 10^{-7}$. The dispersion relation (21) essentially corresponds to the power-law expansion so for practical purposes (due to smallness of expansion parameter E/E_{QG}) only the lowest terms of the expansion are relevant. The relation (21) leads to a hamiltonian

$$\mathcal{H} = \sqrt{p^2c^2[1 + \epsilon \left(\frac{E}{\xi_n E_{QG}} \right)^n]} \tag{22}$$

from which time dependent group velocity $v(t) = \frac{\partial \mathcal{H}}{\partial p}$ can be inferred.

Then, the comoving distance travelled by photon to the Earth is

$$r(t) = \int_{t_{emission}}^{t_{detection}} v(t)dt = \int_0^z v(z') \frac{dz'}{H(z')(1+z')} \tag{23}$$

where in the last equation a standard time-redshift parametrization was taken into account. One can expect these considerations having clear cosmological context. The reason for this is simple — the modifications due to LIV theories are really tiny, so one has to look for sources located at cosmological distances (such like quasars or gamma ray bursts) which are far enough to compensate for the smallness of LIV corrections. This means that cosmological background geometry should be taken into account. Expressing group velocity in terms of redshift, we get

$$v(z) \simeq c(1+z)[1 + \epsilon \frac{(n+1)}{2} \left(\frac{E}{\xi_n E_{QG}} \right)^n (1+z)^n]$$

(24)

Time of flight for the photon of energy E is equal to

$$t_{LIV} = \int_0^z [1 + \epsilon \frac{n+1}{2} \left(\frac{E}{\xi_n E_{QG}} \right)^n (1+z')^n] \frac{dz'}{H(z')}$$

(25)

In the first term one easily recognizes the time of flight for photons in standard relativistic cosmology (i.e. without LIV). Due to very small magnitude of LIV corrections it also fairly represents the time of flight for low energy photons. Consequently, the time delay between a low energy and a high energy photon is equal to

$$\Delta t_{LIV} = \frac{n+1}{2} \left(\frac{E}{\xi_n E_{QG}} \right)^n \int_0^z \frac{(1+z')^n dz'}{H(z')}$$

(26)

where we restricted our attention to "infraluminal" motion of high energy photons (i.e. low energy photons arrive earlier to the observer). Generalization to "superluminal" motion is straightforward — time delays become early arrivals. The idea of observational strategy emerging form (26) is simple. One should monitor appropriate (i.e. emitting both low and high energy photons) cosmological source at different energy channels and try to detect this time delay. However there remains an indispensable uncertainty about intrinsic time delays: there is no reason for which low and high energy signal should be emitted simultaneously. The method outlined below, based on strong gravitational lensing allows to get rid of this ambiguity.

Let us now imagine a source at cosmological distance emitting low energy and high (TeV) energy photons which is gravitationally lensed by a foreground galaxy. Let us also assume that LIV type distorted dispersion relation (21) holds. The observer would also notice time delays between images, but this time it would be a combined effect of gravitational lensing and LIV. Therefore it would no longer be achromatic.

It is easy to calculate this by using the fictitious "LIV comoving distance" $r_{LIV}(z)$, namely:

$$\Delta t_{LIV,SIS} = \frac{8\pi}{H_0} \widetilde{r}_{LIV}(z_l) \beta \frac{\sigma^2}{c^2}$$

(27)

where:

$$\widetilde{r}_{LIV}(z_l) = \widetilde{r}_l + H_0 \frac{n+1}{2} \left(\frac{E}{\xi_n E_{QG}} \right)^n \int_0^{z_l} \frac{(1+z')^n dz'}{H(z')}$$

(28)

Now we can assume that observations in low energy would essentially provide time delay between images equal to Δt_{SIS}, whereas monitoring of the same images in high energy (TeV)

channel would provide $\Delta t_{LIV,SIS}$. Restricting attention to the $n = 1$ case (because LIV effects are extremely small), we can see that these two measurements would differ by

$$\Delta t_{LIV,SIS} - \Delta t_{SIS} = \frac{8\pi}{H_0}\beta\frac{\sigma^2}{c^2}\frac{E}{E_{QG}}\int_0^z\frac{(1+z')dz'}{H(z')} \tag{29}$$

Let us make an estimate for the above LIV effect taking a real strong lensing system. To be specific, we will assume a flat FRW model with $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$ — the so called concordance model. Strong lensing system HST 14176+5226 can serve as an example. This system was discovered with the Hubble Space Telescope and further confirmed to be a gravitational lens. The lensed source is a quasar at redshift $z_s = 3.4$ whereas the lens is an elliptical galaxy having redshift $z_l = 0.809$. The lens model best fitted to the observed images gives the Einstein radius $\theta_E = 1''.489$ and $\beta = 0''.13$. Optical spectroscopy of the lensing galaxy in HST 14176+5226 system provided measurements of the velocity dispersion of $\sigma = 290 \pm 8\ km/s$ in the lensing galaxy. Substituting these data to (29) gives $\Delta t_{LIV,SIS} - \Delta t_{SIS}$ equal to $3.7 \times 10^{-9}\ s$ for 5 TeV photons and $1.5 \times 10^{-8}\ s$ for 20 TeV ones.

It would also be interesting to ask how the LIV effects might modify image configurations. It could be suspected that they might do so since from the Fermat's principle perspective images are located at stationary points of the wavefront travel time functional (given by equation (10)). Therefore since LIV modifies time of flight in an energy dependent way (due to modified dispersion relation) then one expects the images seen at different energies located at different positions. It is easy to see that for the SIS lens (generalizations to other mass profiles are also rather straightforward) the difference between Einstein radii for high and low energy photons $\Delta\theta_{E,LIV} := \theta_{E,LIV} - \theta_E$ would be given by formula:

$$\Delta\theta_{E,LIV} = \theta_E\frac{E}{E_{QG}}\left(\frac{I^{(1)}(z_l, z_s)}{\widetilde{r}(z_l, z_s)} - \frac{I^{(1)}(z_s)}{\widetilde{r}(z_s)}\right) \tag{30}$$

where: $I^{(1)}(z_1, z_2) := \int_{z_1}^{z_2}\frac{(1+z')dz'}{H(z')}$. For realistic lens configurations like HST 14176+5226 this would give negligibly small corrections of order 10^{-16} arc sec. Hence even if LIV were operating this would not be able to change macro-images position in a detectable way. However it cannot be excluded that such minute differences could become relevant while studying caustic crossing (Schneider et al., 2006) possibly leading to different magnification patterns due to microlensing at different energies.

One may ask if appropriate lensing systems (i.e. having sources emitting both low and high energy photons) exist. It is an observational fact that very high energy emission ($E > 100\ GeV$) has been detected from over a dozen of blazars which have similar nature with quasars. Quasars, on the other hand are the sources in all known strong lensing systems. It is a matter of coordinating strong lensing surveys with experiments in high energy astrophysics, such like AGILE, GLAST or MAGIC experiments and the future will certainly bring the discovery of lensed high energy source.

In conclusion, one should keep in mind the possibility of testing LIV effects by monitoring time delays between images of gravitationally lensed quasars in low and high energy channels. In standard theory (General Relativity) the result should be the same — gravitational lensing is essentially achromatic. On the other hand in the presence of LIV effects time delays loose this property — high energy photons should come at different times comparing with low energy ones. Therefore time delays between images should be

different at different energies (e.g. optical or gamma-rays and TeV photons). Because this method is differential in nature, it gets rid of the assumptions about intrinsic time delays of signals at different energies. In fact time delays between images at different energies could be established in different experiments (at unrelated observing sessions) performed on given lensing system. The only demand is that they are accurate enough (done with a sufficient temporal resolution). However, light curves of gamma-ray bursts are already sampled with mili-second resolution and AGILE experiment went down to micro-seconds.

7. Conclusions

Strong gravitational lensing as an effect rooted deeply in General Relativity has great potential in constraining many aspects of gravitational physics. First of all it is useful in studies of dark matter. It stems from the fact that gravitational lensing is sensitive to mass distribution regardless of its nature (whether they are baryonic or not). This is already a rich field being currently explored both theoretically and observationally.

In all known strong lensing systems producing multiple images, the population of sources is of cosmological nature (quasars or distant bright galaxies). In light of recent progress in modeling lensing galaxies, and considerable enrichment of observational data with reliable spectroscopic measurements allowing for determination of redshifts and central velocity dispersions, the new possibility opens up to use well studied strong lensing systems for constraining cosmological model parameters. Although in the past, there was certain scepticism about this technique it is currently proving its effectiveness and in the future – having in mind development of ongoing and planned lens surveys – it will eventually evolve into a competitive technique for cosmography. This is very important, because of the dark energy problem (i.e. the puzzle of presently accelerating Universe). Currently the only empirical way to address the issue of dark energy is by refining the cosmography.

At last, since the light deflection angle by a massive bodies, which is the corner stone of gravitational lensing, is usually calculated in General Relativistic framework, strong lensing can also be used for investigating alternative theories of gravity. It could be also helpful in the context of searching the energy dependent time delays as predicted by Lorentz Invariance Violating theories. Therefore, one may hope for a bright future of strong gravitational lensing is bright and one can expect that it will bring us many interesting results especially when observing campaigns of strong lenses (besides optical or radio regimes) are extended into very high energy range.

8. References

Amanullah, R., et al. (2010). *Astrophys.J.*, Vol.(716):712

Amelino-Camelia, G., Ellis, J.R., Mavromatos, N.E., Nanopoulos, D.V. & Sarkar, S. (1998). *Nature*, Vol(393):763

Amelino-Camelia, G. & Piran, T. (2001). *Phys.Rev. D*, Vol.(64):763

Amendola, L., Finelli, F., Burigana, C. & Carturan, D. (2003). *J. Cosmol. Astropart. Phys.*, Vol.(07):005

Angus, G. W., Shan, H. Y., Zhao, H.-S. & Famaey, B. (2006). *Astrophys. J.*, Vol.(654):L13

Bassett, B.A. & Kunz, M. (2004). *Phys. Rev. D*, Vol.(69): 101305

Bekenstein, J. D. (2004). *Phys. Rev.D*, Vol.(70):083509

Bekenstein, J. D. (2010). Alternatives to dark matter: Modified gravity as an alternative to dark matter, In: *Particle Dark Matter: Observations, Models and Searches*, Bertone, G., (Ed.), 95-114, Cambridge U. Press, Cambridge

Biesiada, M., Godłowski, W. & Szydłowski, M. (2005). *Astrophys. J.*, Vol.(622):28

Biesiada, M. (2006). *Phys. Rev. D*, Vol.(73):023006

Biesiada, M. & Piórkowska, A. (2009). *Mon. Not. R. Astron. Soc.*, Vol.(396):946

Biesiada, M., Piórkowska, A. & Malec, B. (2010). *Mon. Not. R. Astron. Soc.*, Vol.(406):1055

Biesiada, M., Malec, B. & Piórkowska A. (2011). Dark energy constraints from joint analysis of standard rulers and standard candles. *Res. Astron. Astrophys.*, Vol(11): in print

Bolton, A.S., Burles, S., Koopmans, L.V.E., Treu, T. & Moustakas, L.A. (2006). *Astrophys. J.*, Vol.(638):703

Bradač, M., Allen, S. W., Treu, T., Ebeling, H., Massey, R., Morris, R. G., von der Linden, A. & Applegate, D. (2008). Revealing the properties of dark matter in the merging cluster MACS J0025.4-1222. *Ap J.*, Vol.(676):959

Buchert, T. (2001). *Gen. Relativ. Grav.*, Vol.(33):1381

Caldwell, R.R. & Kamionkowski M. (2009). The Physics of Cosmic Acceleration. *Ann.Rev.Nucl.Part.Sci.*, Vol.(59):397

Chae, K.H. (2007). *Astrophys. J.*, Vol(658):L71

Chevallier, M., & Polarski, D. (2001). *Int. J. Mod.Phys.D*, Vol.(10):213

Chiu, M. C., Ko, C. M. & Tian, Y. (2006). *Astrophys. J.*, Vol.(636):565

Clowe, D., Bradač, M., Gonzalez, A. H., Markevitch, M., Randall, S. W., Jones, C. & Zaritsky, D. (2006). *Astrophys. J.*, Vol.(648):L109

Coles, J. (2008). *Astrophys. J.*, Vol.(679):17

Cyburt, R.H. (2004). *Phys. Rev. D*, Vol.(70):023505

Dunkley, J., Komatsu, E., Nolta, M.R. et al. (2009). *Astrophys.J. Suppl.*, Vol(180):306

Dvali, G., Gabadadze, G. & Porrati, M. (2000). *Phys. Lett.B*, Vol.(485):208

Eisenstein, D.J. & Hu, W. (1999). *Astrophys.J.*, Vol.(511):5

Ellis, J.R., Mavromatos, N.E., Nanopoulos, D.V., Sakharov, A.S. & Sarkisyan, E.K.G. (2006). *Astropart.Phys.*, Vol.(25):402

Feldman, H., et al. (2003). *Astrophys.J.*, Vol.(596):L131

Freedman, W.L., et al. (2001). Final results from the HST Key Project to measure the Hubble Constant. *Astrophys. J.*, Vol.(553):47

Futamase, T. & Yoshida, S. (2001). *Prog. Theor. Phys.*, Vol.(105):887

Gilmore, J. & Natarayan, P. (2009). *Mon. Not. R. Astron. Soc.*, Vol.(396):354

Grillo, C., Lombardi, M. & Bertin, G. (2008). *Astron. Astrophys.*, Vol.(477):397

Hinshaw, G., et al. (2009). *Astrophys.J. Suppl.*, Vol.(180):225

Jullo, E., Natarajan, P., Kneib, J., D'Aloisio, A., Limousin, M., Richard, J., Schimd, C. (2010). *Science*, Vol.(329):924

Komatsu, E., et al. (2009). *Astrophys.J. Suppl.*, Vol.(180):330

Koopmans, L.V.E., Treu, T., Bolton, A.S., Burles, S. & Moustakas, L.A. (2006). *Astrohys.J.*, Vol.(649):599

Koopmans, L.V.E. et al. (2009). *Astrophys.J.*, Vol.(703): L51

Lazkoz, R., Nesseris, S. & Perivolaropoulos, L. (2008). *J. Cosmol. Astropart. Phys.*, Vol.(0807):012

Linder, E.V. (2003). *Phys. Rev.D*, Vol.(68):083503

Linder E.V., Roberts G. (2008). *J. Cosmol. Astropart. Phys.*, Vol.(0806):004

Meneghetti, M., Bartelmann, M., Dolag, K., Moscardini, L., Perrotta, F., Baccigalupi, C. & Tormen, G. (2005). *Astron. Astrophys.*, Vol.(442):413

Milgrom, M. (1983). *Astrophys. J.*, Vol.(270):365

Milgrom, M. (2008). *New Astronomy Reviews*, Vol.(51):906

Mohayaee, R. & Tully, B. (2005). *Astrophys.J.*, Vol.(635):L113

Mortlock, D. J. & Turner, E. L. (2001). *Mon. Not. R. Astron. Soc.*, Vo.(327):557

Oguri, M. (2007). *New J.Phys.*, Vol.(9):442

Oguri, M., Inada, N., Strauss, M.A., Kochanek, C.S. & Richards, G.T., et al. (2008.) *Astron. J.* Vol.(135):512

Paczyński, B. (1987). *Nature*, Vol.(325):572

Paczyński, B. & Górski, K. (1981). *Astrophys. J.*, Vol.(248):L101

Percival, W.J., et al. (2007). *Astrophys.J.*, Vol.(657):51

Perlmutter, S., et al. (1999). *Astrophys.J.*, Vol.(517):565

Refsdal, S. (1964). *Mon. Not. R. Astron. Soc.*, Vol.(128):307

Riess, A.G., et al. (1998). *Astron. J.*, Vol.(116):1009

Sandage, A. (1970). *Physics Today*, Vol.(23):34

Sanders, R. H. & McGaugh, S. S. (2002). *Ann. Rev. Astron. Astrophys.*, Vol.(40):263

Sereno, M. (2002). *Astron. Astrophys.*, Vol.(393):757

Schneider, P., Ehlers, J. & Falco, E. E. (1992). *Gravitational Lenses*, Springer Verlag, Berlin.

Schneider, P., Kochanek, C.P. & Wambsganss, J. (2006). Gravitational Lensing: Strong, Weak & Micro. In, *Proceedings of the 33rd Saas-Fee Advanced Course*, Meylan, G., Jetzer, P., North, P., (Ed.), Springer-Verlag, Heidelberg

Spergel, D.N., et al. (2003). *Astrophys.J. Suppl.*, Vol.(148):175

Taylor, J.H., Wolszczan, A., Damour, T. & Weisberg, J.M. (1992). *Nature*, Vol.(355):132

Treu, T., Koopmans, L.V.E., Bolton, A.S., Burles, S. & Moustakas, L.A. (2006). *Astrophys.J.*, Vol.(650):1219

Uzan, J.P., Aghanim, N. & Mellier, Y. (2004). *Phys.Rev. D*, Vol.(70):083533

Weinberg, S. (1989). *Rev. Mod. Phys.*, Vol.(61):1

Wiltshire, D. (2007). *Phys. Rev. Lett.*, Vol.(99):251101

Xu, L. & Wang, Y. (2010). *Phys Rev D* Vol.(82):043503

Wu, P., & Yu, H. (2007). *Astrophys. J.*, Vol.(658):663

Zhao, H.-S., Bacon, D. J., Taylor, A. N. & Horne, K. (2006). *Mon. Not. R. Astron. Soc.*, Vol.(368):171

Part 3

Theoretical Investigations

The Dirac Field at the Future Conformal Singularity

Michael Ibison

Institute for Advanced Studies at Austin,
USA

1. Introduction

In the flat space Friedmann-Robertson-Walker (FRW) Cosmology expressed in the coordinate system with line element $ds^2 = d\tau^2 - a^2(\tau)d\mathbf{x}^2$, τ is the proper time of a 'fundamental observer' equipped with a suitably-defined 'laboratory clock' at rest with respect to the frame established by the Cosmic Microwave Background. Measured according to this time The Universe will expand forever, asymptotically approaching an exponential growth $a(\tau) \to \exp(H\tau)$, where H is the Hubble parameter. Descriptions and explanations of Cosmology commonly adopt this time implicitly - though GR is coordinate independent of course. For example the cosmological (as opposed to Doppler) red-shift of distant galaxies is usually described as due to the effect of the expansion of space during the time light – or a photon – is in flight from the distant star on its way to Earth. Unless properly qualified such explanations give the false impression that a physical statement is being made about the effect of expansion on light, without recognizing that the explanation makes use of a projection of the underlying physics onto a particular coordinate system. In this instance the physical – coordinate independent – essence of the phenomenon depends on the differential evolution of electron mass and photon energy, and not on the photon energy by itself. So for example a perfectly good alternative explanation for the same observation is that the Lyman Alpha lines of local hydrogen, i.e. at the time of reception, are more separated energetically (bluer) then those at the time of emission, with light having suffered no red-shift at all whilst in flight. This alternative picture comes from projecting the physical process onto the conformal coordinate system having line element $ds^2 = a^2(t)(dt^2 - d\mathbf{x}^2)$. In contrast with the 'traditional' coordinate system, in the conformal system $(-g)^{1/2}$ times the energy density of matter increases in proportion to the scale, giving rise, effectively, to an increase in the rest mass of a free electron. Of course there are infinitely many other coordinate systems. But this one example serves to illustrate the difference between a process and its projection.

One does not expect to find any new physics simply by changing the coordinate system; a coordinate transformation in GR is analogous to a coordinate rotation in Euclidean geometry. Yet there is a concern. The FRW system has no future boundary, whereas the conformal scale factor is singular at a finite conformal time; the exponential expansion written conformally has asymptote $a(t) \to (1 - Ht)^{-1}$. One might wonder if perhaps this is a coordinate singularity with no physical consequence and can be ignored? As implied above (though with some qualifications) EM radiation is unaffected by the expansion expressed in

the conformal system – perhaps it is unaffected also by the singularity? And how is fermionic matter affected by the singularity?

Penrose, and Friedrich (Friedrich, H., 2002; Penrose, R., 1963) have considered in detail the consequences of the conformal singularity for gravitational radiation in particular. The focus here is on the consequences for the *Dirac field*. We will show below that Dirac's equation in conformal spacetime with conformal scale factor $f(x)$ can be written

$$\left(\gamma^{\mu}\left(-i\partial_{\mu}+eA_{\mu}(x)\right)+f(x)m\right)\psi(x)=0\,. \tag{1}$$

Of interest is the manner in which $f(x)$ introduces a dependency on absolute coordinates and so changes the way the discrete symmetry operations affect the equation as a whole. A conformal factor due to cosmological expansion has a different status than the vector potential because it is a fixed background affecting all particles everywhere. Arguably it should be considered as hard-wired into the Dirac equation in the same manner that the (constant) mass term is ordinarily considered a fixed feature. In fact since it multiplies the mass one can take the position that a *coordinate dependent mass* is a universal property of the Dirac equation. Obviously this point of view pertains specifically to the equation expressed in conformal coordinates.

We are interested in the ways the conformal factor influences behavior: 1) On the local modification of discrete symmetries, 2) On the relationship between pre and post-singularity wavefunctions, 3) The boundary conditions on the wavefunction and / or the topologies of Cosmological spacetime required in order that the wavefunction behave nicely *through* the singularity. These topics are covered in the subsequent sections as follows. Section 2 reviews the symmetry-breaking effects of the conformal metric on the discrete symmetries normally present in Minkowski spacetime. The latter are reviewed in Appendix A. Section 3 looks at the behavior of the conformal scale factor near and through the singularity as determined by the Friedmann equation for a conformally-expressed metric in the flat space Robertson-Walker spacetime. Section 4 gives a very brief review of EM in conformal spacetime. The affect of the conformal metric on the Dirac equation in general, and the wavefunction in particular are covered in Sections 5 and 6 respectively. Those findings are then applied in Section 7 to the particular case of a conformal representation of the de Sitter spacetime. Those results are further specialized in Section 8 in an analysis of the wavefunction near and through the singularity. (With some qualifications, all vacuum-dominated Robertson-Walker spacetimes asymptote to the de Sitter evolution). Section 9 outlines the alternatives for peaceful coexistence between the Friedmann equation and the Dirac wavefunction under the presumption that the post-singularity universe is not a redundant copy of the pre-singularity universe.

2. Inversion symmetries

2.1 Systematic symmetries

An inversion operation may be a symmetry of the whole of *system* of physical interactions if applied universally. By universally we mean here not just over all space and time, but to all particles. It is easy to see for example that the system of QED must be invariant under charge conjugation. QED is invariant also under parity and time reversals, independently. Since the system as a whole is invariant, it follows that an inversion applied universally to a physically legal universe of particles and their interactions generates another universe, legal

under the rules of QED. It is different matter however to determine if this implies a symmetry (point symmetry or local symmetry) of the Dirac wavefunction, either free, or in the presence of an interaction.

2.2 Point symmetry and local symmetry in the dirac equation

A free particle solution in Minkowski spacetime $\psi(t,\mathbf{x})$ i.e. obeying (1) with $A_\mu(x) = 0$, $f(x) = 1$,

$$\left(-i\gamma^\mu \partial_\mu + m\right)\psi(x) = 0 \tag{2}$$

turns out to have symmetric partners associated with time reversal and parity inversion. Having fixed a coordinate system with a particular origin, technically, the replacement $\mathbf{x} \to \mathbf{x}' = -\mathbf{x}$ has the specific meaning of inverting the coordinates through the spatial origin $\mathbf{x} = \mathbf{0}$. Therefore the presence of a symmetry under this operation implies that for every solution $\psi(t,\mathbf{x})$ there exists another solution $\psi'(t,\mathbf{x})$ of the same equation that is somehow related to the solution at $\psi(t,-\mathbf{x})$. For example the relationship may be of the form $\psi'(t,\mathbf{x}) = U\psi(t,-\mathbf{x})$, where U is some fixed 4x4 matrix. If the origin remains where it was before the inversion this relates solutions that are spatially separated, depending on their distance from the origin. That is, $\psi'(t,\mathbf{x}) = U\psi(t,-\mathbf{x})$ is a *point symmetry*. To be concrete, if the origin were at the center of the Milky way, then the symmetry implies that for every Earth-based particle with wavefunction $\psi(t,\mathbf{x})$ there may be another particle with wavefunction $U\psi(t,-\mathbf{x})$ located on the 'other side' of the Milky Way at a distance of about 100,000 light years from Earth. Usually though this is not what is meant by parity inversion symmetry in the context of the Dirac equation. Instead it is understood that there is a freedom to combine the replacement $\mathbf{x} \to \mathbf{x}' = -\mathbf{x}$ with an arbitrary translation to bring the symmetric partner to the same location as the original. This is possible because (2) is translation invariant. In practice this freedom is used to move the origin to the location of the original particle before applying the operation $\mathbf{x} \to \mathbf{x}' = -\mathbf{x}$, with the outcome that inversion can be treated as an entirely local operation. The combination of translation invariance with parity inversion invariance converts the point symmetry to a *local symmetry*.[1]

When the vector potential is absent one usually expects to see the local version of the time and parity symmetries. Generally, the effect of a (universal) conformal scale factor is to destroy the translation invariance to some degree in the same manner as would the presence of a vector potential. With translational invariance gone only the point symmetry versions remain, these now to be interpreted with respect to the absolute coordinate system established by the metric and, in particular, the conformal singularity. Consequently in the following we will be interested in how the conformal factor affects the local symmetry, and additionally in the consequences of its replacement with a point symmetry.

3. The Friedmann equation in conformal spacetime

3.1 Conformal forms of the scale factor

We consider in parallel the two forms of Cosmological scale factor

$$f(x) \in \left\{ a(t), a\left(t / x^2\right) / x^2 \right\} \tag{3}$$

---

[1]This point of view is applicable only to a single particle, which is our sole interest here.

where a is an arbitrary function common to both cases determined from solution of the Friedmann equation. The notations used here are $x = \{x^\mu\} = (t,\mathbf{x})$, $r = |\mathbf{x}|$. We note in passing that the first of these admits the alternative form

$$f(x) = \frac{1}{t} b\left(\frac{t}{x^2}\right) \tag{4}$$

where $b(z) = za(z)$. Each of (3) is associated with a line-element

$$ds^2 = f^2(x)\,dx^2 \tag{5}$$

which is why they are called conformal. The coordinates used for these definitions are not the same though we have used the same symbols - they are related by the coordinate transformation

$$t \to t/x^2, \quad r \to r/x^2 \tag{6}$$

with the angle variables left unchanged. Note that the transformation is symmetric in that it can be applied to take the first of (3) to the second and vice-versa. Because a coordinate transformation exists between these two forms they are equivalent from a GR point of view, though they may *imply* different topologies. When expressed as a Robertson-Walker spacetime the first of these has zero spatial curvature. That is, it can be written in the form

$$ds^2 = d\tau^2 - a'^2(\tau)\,d\mathbf{x}^2 \tag{7}$$

where $d\tau = a(t)\,dt$ and $a'(\tau) = a(t)$ is a (new) arbitrary function. Since the two systems are related by a transformation the second of (3) can also be put into the form (7) and therefore also has zero spatial curvature in that context. (Spatial curvature is not a coordinate-independent quality of a metric. It is sufficient to note that all Robertson-Walker spacetimes can be expressed conformally, removing therefore the spatial curvature from the $K = \pm 1$ spacetimes - see (Ibison, M., 2007).) In the form $f(x) = a(t)$ the (hyper) surfaces of cosmological simultaneity are the hyper-planes $t = $ constant. In the case $f(x) = a(t/x^2)/x^2$ then $t = kx^2$ for some constant k and then

$$\left(t - \frac{1}{2k}\right)^2 - r^2 = \left(\frac{1}{2k}\right)^2. \tag{8}$$

Hence the surfaces of simultaneity are paraboloids, including the 'final surface' that is the conformal singularity. Despite superficial appearances therefore, both forms in (3) admit three independent translational isometries.

3.2 Two branches
The evolution of the scale factor is decided by the Friedmann equation plus equations of state for the various contributions.[2] In conformal coordinates this is

---

[2]Either the equations of state or the second Friedmann equation involving the pressure.

$$\frac{1}{H^2}\left(\frac{da}{dt}\right)^2 = \Omega_{EM} + \Omega_m a + \Omega_\Lambda a^4 \,. \tag{9}$$

Each of the Ω is an energy density normalized so that their sum is unity; Λ denotes the vacuum contribution. Present estimates are (Nakamura, K. & et al, 2010)

$$H^{-1} = 9.78 \,/\, .72 = 13.6 \quad \text{Gyr}$$
$$\Omega_\Lambda = 0.74, \quad \Omega_m = 0.256, \quad \Omega_{EM} = 4.76 \times 10^{-5} \,. \tag{10}$$

Let us write the Friedmann equation as

$$\frac{1}{H^2}\left(\frac{da}{dt}\right)^2 = f(a) \tag{11}$$

where $f(a)$ is a dimensionless function of the scale factor. Upon integration one has

$$t_+(a) = \frac{1}{H}\int^a \frac{da'}{\left|\sqrt{f(a')}\right|} \quad \text{and} \quad t_-(a) = -\frac{1}{H}\int^a \frac{da'}{\left|\sqrt{f(a')}\right|} \tag{12}$$

I.E. t has two (single-valued) branches $t_+(a)$ and $t_-(a)$ which are true functions. (The square root operation is discussed in more detail below.) Since the integrand is always positive or zero the $t_\pm(a)$ are both monotonic with a. Therefore each function is invertible, and in each of these the scale factor can be expressed as a function of time. Let us write the Taylor-Laurent series expansion of $f(a)$ about $a = 0$ as

$$f(a) = c_m a^m + c_{m+1} a^{m+1} + \ldots + c_{n-1} a^{n-1} + c_n a^n; \quad n > m \tag{13}$$

and set $a = 1$ at the present – finite – time. Then the integrals in (12) converge to a finite future t as $a \to \infty$ provided $n > 2$. In that case there is a singularity in the scale factor in finite conformal time. Similarly the integrals converge to a finite t as $a \to 0_+$ provided $m \geq 0$. In that case the Big Bang occurred at a finite conformal time in the past. In (9) $m = 2$, $n = 4$, and both of these conditions are met; the universe has a finite conformal duration. Defining

$$\phi(a) := \frac{1}{H}\int_\infty^a \frac{da'}{\left|\sqrt{f(a')}\right|} \tag{14}$$

the solutions (12) can be written

$$t_+(a) = \phi(a) + t_+(\infty) \quad \text{and} \quad t_-(a) = -\phi(a) + t_-(\infty) \tag{15}$$

where we have chosen to set whatever initial conditions we intend to apply at the conformal singularity (which will now be presumed to exist). Let us set the clocks to zero there - $t_+(\infty) = t_-(\infty) = 0$ - rather than at the Big Bang. With this, and inverting (15), the inversion of the two branches gives

$$a_+(t) = \phi^{-1}(t) \quad \text{and} \quad a_-(t) = \phi^{-1}(-t) \,. \tag{16}$$

3.3 Parity of the scale factor

Usually there would be no motivation to entertain solutions of the Friedmann equation for negative values of the scale factor. But since the conformal boundary occurs at a finite time and the scale changes sign thereafter, this would appear to be an oversight. We will see below that the energy of Dirac matter changes sign through the boundary, with the result that the product $\Omega_m a(t)$ is always positive. Consequently (9) is more accurately written

$$\frac{1}{H^2}\left(\frac{da}{dt}\right)^2 = \Omega_{EM} + |\Omega_m a| + \Omega_\Lambda a^4 \tag{17}$$

Now the right hand side has even parity. Consequently $f^{1/2}(a) = f^{1/2}(-a)$ so that the even parity survives the square root operation. No questions arise as to the meaning of the square root operation here and in (14) because $f(a)^{1/2}$ is positive for all real a.

Since the integrand has even parity $\phi(a)$ is odd. The two branches in (16) then appear as in Figure 1, with scaling of a so that $a_+(0) = - a_-(0) = 1$. The branch $a_+(t)$ is shown in solid blue and the $a_-(t)$ branch as a red dashed curve. A solution valid through the singularity now has odd parity (everywhere) - consistent with $a \sim 1/t$. The Friedmann equation 'predicts' a post singularity universe that is a mirror image of (our) pre-singularity universe though with $a \to -a$. For the Friedmann equation to remain valid the image must be a legal – dynamically feasible – copy of the original pre-singularity universe.
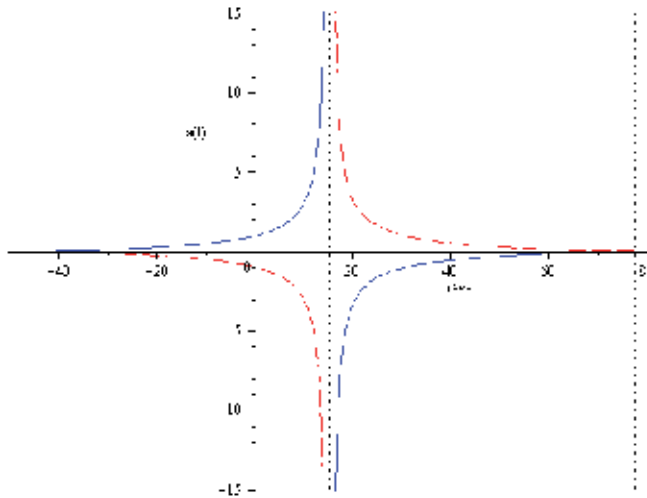
Fig. 1. Plot of both branches of the scale factor on both sides of the conformal singularity; the solid curves are a single branch and the broken curves are a single branch. Here the origin of the time coordinate has been chosen so that $a(0) = 1$.

3.4 Singularities and asymptotic behavior

The time to the singularity found by integrating (9) numerically is

$$t(a=\infty) - t(a=1) = \frac{1}{H}\int_1^\infty \frac{da}{\sqrt{\Omega_{EM} + \Omega_m a + \Omega_\Lambda a^4}} = 1.12 / H = 15.2 \ \text{Gyr} \tag{18}$$

which is marked in Figure 1 by a dotted line. The conformal time elapsed since the Big Bang – the age - is

$$t(a=1)-t(a=0)=\frac{1}{H}\int_0^1 \frac{da}{\sqrt{\Omega_{EM}+\Omega_m a+\Omega_\Lambda a^4}}=3.47\,/\,H=47.2\ \text{Gyr}\,. \tag{19}$$

The conformal interval from Big Bang to future singularity is therefore 62.4 GYr, which we will call the duration. Focusing on the branch $a_+(t)$, the Big Bang was at time $t = -47.2$ GYr ago. The scale factor expands from 0 to positive infinity at 15.2 Gyr in the future. The scale factor then changes sign and proceeds to diminish in magnitude from negative infinity to 0, which it reaches at time 15.2 + 62.4 = 77.6 Gyr.

Near the future singularity the vacuum-dominated asymptotic behavior is that of a simple pole

$$a(t)\rightarrow (1-kt)^{-1}\,. \tag{20}$$

Therefore (20) with $k = 1/15.2$ Gyr is an approximation to the remaining evolution which ignores the non-asymptotic behavior. In the following we will exploit the fact that (9) is invariant under time translations to move the time of the future singularity to $t = 0$ with the result that the present time is *negative* 15.2 Gyr. The evolution near the singularity is then like $a \sim 1/t$ and is odd about the new origin. This has the advantage that if the transformation (6) is applied to this system (with this origin) then the conformal singularity occurs at the same time for both systems defined in (3) i.e. at $t = 0$. Further, the asymptotic behavior $a \sim 1/t$ corresponds in the system with $f(x) = a(t/x^2)/x^2$ to $f(x) \sim 1/t$ also, so both systems have the same asymptotic behavior. After all this, near the singularity we can then ignore the differences because the two forms are essentially the same. We should point out however that these are not the only forms of conformal factor that represent de Sitter spacetime. The curved-space Robertson-Walker spacetime for example has a different de Sitter asymptote (Lasenby, 2002; Lasenby, A. & Doran, C., 2005) which is not covered by the analysis here, and would be interesting to investigate in this context.

4. EM in conformal spacetime

The EM action in curved spacetime is

$$I=-\int d^4 x\sqrt{-g}\left(\frac{1}{4}F_{ab}F^{ab}+A_a j^a\right). \tag{21}$$

In the particular case of conformal spacetime it is useful to re-write this as

$$I=-\int d^4 x\left(\frac{1}{4}F_{ab}F_{bd}\eta^{ac}\eta^{bd}+A_a\sqrt{-g}\,j^a\right). \tag{22}$$

The covariant divergence of the current must vanish

$$j^a_{\ ;a}=0\Rightarrow \partial_a\left(\sqrt{-g}\,j^a\right)=0\,. \tag{23}$$

Apart from this, the equations are the same as for Minkowski spacetime. Therefore it is convenient to define

$$\bar{j}^a = \sqrt{-g}\, j^a \tag{24}$$

whose ordinary divergence must now vanish. With this, the action (22) becomes

$$I = -\int d^4x \left(\frac{1}{4} F_{ab} F_{bd} \eta^{ac} \eta^{bd} + A_a \bar{j}_b \eta^{ab} \right) \tag{25}$$

and now the scale factor has been eliminated. It follows that variation of the covariant potentials in (25) must give the Maxwell equations as if in Minkowski spacetime:

$$\partial^2 A_a - \partial_a (\partial \circ A) = \bar{j}_a \,. \tag{26}$$

The Minkowski spacetime Lorenz gauge

$$\partial \circ A \equiv \eta^{ab} \partial_a A_b = 0 \tag{27}$$

then leads to

$$\partial^2 A_a = \bar{j}_a \,. \tag{28}$$

Appendix A gives the discrete inversion symmetries in Minkowski spacetime. It will be necessary for subsequent discussions to know how these are affected in going to conformal spacetime. Of particular interest is the fate of time reversal symmetry due to the presence of a scale factor that is odd about the conformal singularity. Since the scale factor is absent in the Lorenz gauge, the symmetric partner of $A^\mu(x)$ is $h^\mu_{\ \nu} A^\nu(-\tilde{x})$ just as in Minkowski spacetime, though this is demoted to a point symmetry rather than a local one.

Consider now the consequences of imposing instead the covariant gauge condition

$$A^\mu_{\ ;\mu} = \frac{1}{\sqrt{-g}} \partial_\mu \left(\sqrt{-g}\, A^\mu \right) = 0 \Rightarrow \partial_\mu \left(a^4 A^\mu \right) = 0 \Rightarrow 2\frac{\dot{a}}{a}\phi + \frac{\partial\phi}{\partial t} + \nabla.\mathbf{A} = 0 \tag{29}$$

where the 1 + 3 potentials are components of a covariant vector:

$$\{A_\mu\} = (\phi, \mathbf{A}); \quad A^\mu = g^{\mu\nu} A_\nu \Rightarrow \{a^4 A^\mu\} = a^2 (\phi, \mathbf{A}) \,. \tag{30}$$

Then (26) is replaced by

$$\partial^2 \phi + 2\frac{\dot{a}}{a}\frac{\partial\phi}{\partial t} + 2\left(\frac{\ddot{a}}{a} - \frac{\dot{a}^2}{a^2} \right)\phi = \bar{\rho}, \quad \partial^2 \mathbf{A} = \bar{\mathbf{j}} + 2\frac{\dot{a}}{a}\nabla\phi \tag{31}$$

where we used

$$\{\bar{j}_\mu\} = \left(\bar{\rho}, \bar{\mathbf{j}} \right). \tag{32}$$

Evidently the components of the current vector are not affected by this change and so transform under time reversals as they did before. It follows from the structure of (31) that

the potentials are likewise unaffected and the symmetric partner of $A^\mu(x)$ remains $h^\mu{}_\nu A^\nu(-\tilde{x})$. Hence the column headed $A^\mu(x)$ in Table 1 remains as it was in Table A2. The fate of EM fields in conformal spacetime is discussed in more detail in (Ibison, 2010).

5. The Dirac equation in conformal spacetime

5.1 Discrete symmetries
In going to conformal spacetime from Minkowski spacetime we need only consider the effects of space and time inversions, since charge and mass inversion are unchanged. Mass inversion however will be 'opposed' by any operation that inverts the sign of the conformal factor. Discrete symmetries broken by the Cosmological metric have been discussed by Tomaschitz (Tomaschitz, R., 1994).

5.2 Time reversal
Here we return to (1), restoring the conformal factor in one of the forms (3), and re-examine the effects of the inversions. In both cases the Dirac equation is no longer time-translation invariant due to the cosmological evolution, though local time reversal symmetry remains approximate valid in our era, far from the conformal singularity. Consider for example the rest energy of the Dirac particle computed from (A26). This is now

$$\langle E \rangle = a(t)m \int d^3x \, \psi^\dagger \gamma^0 \psi \tag{33}$$

which is not in general time-independent. Using (20) (which has an origin $t = 0$, $a(0) = 1$ corresponding to 'now') far from the boundary the energy is monotonically increasing at a rate

$$\frac{d}{dt}\langle E \rangle \approx km \int d^3x \, \psi^\dagger \gamma^0 \psi \tag{34}$$

whereas the energy monotonically decreases for a time-reversed particle. This is just the local version of Cosmological red-shift seen from the perspective of the conformal coordinate system. Red-shift is associated in this coordinate system with an increase in rest mass, whilst the EM fields are unaffected by the expansion. Generally we think of this effect as observable only as a result of interactions over great distances. In principle though the effect could be probed locally according to (34), leading to a local determination of the cosmological arrow of time through a broken time reversal symmetry.

In the presence of cosmological expansion exact time reversal symmetry, if it exists, relates two objects either side of the conformal singularity, i.e. as a point symmetry. Both conformal factors in (3) change sign under time reversal but are otherwise unaffected by the inversions. This changes the sign of $mf(x)$ from what it was in the Minkowski case. That sign change can be accommodated in the Dirac equation by replacing ψ with $\gamma^5\psi$, which change is reflected in the rows for T_+ and T_- in Table 1 compared with Table A2.

5.3 Parity inversion
Parity inversion has no effect on the conformal factor $f(x) = a(t)$. Space translation invariance appears to be absent in the case of $f(x) = a(t/x^2)/x^2$. We assume however a freedom to choose a coordinate system with the spatial origin centered at the particle of interest. To be more

precise, we can exploit the translation invariance of the $f(x) = a(t)$ system and then transform using (6) to the system $f(x) = a(t/x^2)/x^2$, which is equivalent to having performed a space-time translation within a paraboloidal surface of simultaneity. From this perspective parity inversion remains a local symmetry in both coordinate systems, as reflected by the entries in Table 1.

5.4 Summary

The discrete symmetries of Minkowski spacetime are preserved in a cosmological conformal expansion, though with some modifications. The biggest change is in time-reversal symmetry, which ceases to become locally valid, but retains a point symmetry. The lack of a local time-symmetry manifests as a 'Cosmological arrow of time', associated in particular with recession of distant galaxies in the Hubble flow and with Cosmological red-shift.

	i	$\langle E \rangle$	e	m	x	$mf(x)$	$A^\mu(x)$	constraints on EM coupling for point symmetry to exist	constraints on EM coupling for local symmetry to exist	symmetric partner of $\psi(x)$	Dirac representation
P_+	+	+								$\gamma^0 \psi(\tilde{x})$	$\gamma^0 \psi(\tilde{x})$
P_-	-	-	+	+	$\tilde{x}$	+	$h^\mu_\nu A^\nu(\tilde{x})$	$A^\mu(\tilde{x}) = A^\mu(x)$	$A^\mu(x) \to A^\mu(t)$	$\left(\gamma^0 C \gamma^5 \gamma^0 \psi(\tilde{x})\right)^*$	$i\gamma^2 \gamma^5 \gamma^0 \psi^*(\tilde{x})$
T_+	+	+							no exact local symmetry	$C^{-1} \psi^*(-\tilde{x})$	$\gamma^2 \gamma^0 \psi^*(-\tilde{x})$
T_-	-	-	+	+	$-\tilde{x}$	-	$h^\mu_\nu A^\nu(-\tilde{x})$	$A^\mu(-\tilde{x}) = A^\mu(x)$		$\gamma^0 \gamma^5 \psi(-\tilde{x})$	$\gamma^0 \gamma^5 \psi(-\tilde{x})$
C_+	+	+						A^μ held fixed (not negated)	none	$\left(\gamma^0 C \psi(x)\right)^*$	$-i\gamma^2 \psi^*(x)$
C_-	-	-	-	+	x	+	$-A^\mu(x)$			$\gamma^{0*} C^* \gamma^{5*} \gamma^0 C \psi(x)$	$-\gamma^5 \psi(x)$
M_+	+	-								$-\gamma^5 \psi(x)$	$-\gamma^5 \psi(x)$
M_-	-	+	+	-	x	-	$A^\mu(x)$	none	none	$\left(\gamma^0 C \psi(x)\right)^*$	$-i\gamma^2 \psi^*(x)$

Table 1. Inversions in Conformal Spacetime with Scale Factor $f(x) \sim 1/t$.

6. The Dirac wavefunction in conformal spacetime

6.1 Tetrad formulation of the Dirac action

The Dirac equation in Minkowski spacetime is

$$\left(\gamma^\alpha \left(i\hbar\partial_\alpha - eA_\alpha\right) - m\right)\psi = 0 \tag{35}$$

where α is a Lorentz index. In curved spacetime the effects of gravitation can be accounted for with the replacement

$$\begin{aligned}
\partial_\alpha &\to \partial_\mu + \Gamma_\mu(x) \\
\gamma^\alpha &\to \gamma^\mu(x) = V_\alpha{}^\mu(x)\gamma^\alpha
\end{aligned}$$

(36)

where the argument x for the gamma-matrix signifies a different object from the Minkowski spacetime matrix, $g^\mu(x) \neq g^\mu$ and where $V_\alpha{}^\mu$ is a tetrad satisfying

$$g_{\mu\nu}(x) = V^\alpha{}_\mu(x)V^\beta{}_\nu(x)\eta_{\alpha\beta}$$

and therefore

$$g^{\mu\nu}(x) = V^{\alpha\mu}(x)V^{\beta\nu}(x)\eta_{\alpha\beta} = V_\alpha{}^\mu(x)V_\beta{}^\nu(x)\eta^{\alpha\beta}$$

Γ_μ is the spin connection defined by

$$\sigma\left[\Gamma_\nu(x), \gamma^\mu(x)\right] = \partial_\nu \gamma^\mu(x) + \Gamma^\mu_{\nu\rho}\gamma^\rho(x)$$

(37)

where σ is introduced to compare different published results (see historical note below). In conformal spacetimes

$$g_{\mu\nu}(x) = f^2(x)\eta_{\mu\nu}\,.$$

The simplest choice is

$$V^\alpha{}_\mu(x) = f(x)\delta^\alpha{}_\mu \Rightarrow V_\alpha{}^\mu(x) = \delta_\alpha{}^\mu / f(x) \Rightarrow \gamma^\mu(x) = \gamma^\mu / f(x)\,.$$

(38)

Then the spin connection equation is

$$\sigma\left[\Gamma_\nu(x), \gamma^\mu\right]f(x) = -\gamma^\mu \partial_\nu f(x) + \Gamma^\mu_{\nu\rho}\gamma^\rho$$

(39)

where the affine connection for the conformal metric is

$$\Gamma^\mu_{\nu\rho} = \left(\delta^\mu_\rho \partial_\nu + \delta^\mu_\nu \partial_\rho - \eta_{\nu\rho}\partial^\mu\right)f(x)\,.$$

(40)

Putting this into (39) gives

$$\sigma\left[\Gamma_\nu(x), \gamma^\mu\right] = \left(\delta^\mu_\nu \not{\partial} - \gamma_\nu \partial^\mu\right)\phi; \quad \phi = \log f(x)$$

(41)

whose solution is

$$\Gamma_\nu(x) = -\frac{\sigma}{2}\gamma_\nu \not{\partial}\phi$$

(42)

(plus an arbitrary function times a constant matrix). Bearing in mind (36), one has

$$\gamma^\alpha \partial_\alpha \psi = \not{\partial}\psi \to \frac{1}{f}\gamma^\mu\left(\partial_\mu - \frac{\sigma}{2}\gamma_\mu \not{\partial}\phi\right)\psi = \frac{1}{f}\left(\not{\partial} - 2\sigma\not{\partial}\phi\right)\psi$$

$$\Rightarrow \not{\partial}\psi \to f^{2\sigma-1}\not{\partial}\left(\frac{\psi}{f^{2\sigma}}\right) \tag{43}$$

and so the Dirac equation in conformal spacetime is

$$i\hbar f^{2\sigma-1}\not{\partial}\left(\frac{\psi}{f^{2\sigma}}\right) - \frac{e}{f}\not{A}\psi - m\psi = 0 . \tag{44}$$

Defining the normalized wavefunction $\tilde{\psi} = \psi / f^{2\sigma}$, and making the dependencies explicit, this becomes

$$\left(\gamma^\alpha\left(i\hbar\partial_\alpha - eA_\alpha(x)\right) - f(x)m\right)\tilde{\psi}(x) = 0 . \tag{45}$$

6.2 Current conservation

Note that the ordinary divergence of the current vanishes

$$\tilde{j}^\alpha = e\bar{\tilde{\psi}}\gamma^\alpha\tilde{\psi}; \quad \partial_\alpha \tilde{j}^\alpha = 0 \tag{46}$$

and therefore the charge is conserved in Minkowski space. Of course one can revert at any time to the un-normalized wavefunction using $\psi = f^{2\sigma}\tilde{\psi}$, for which the current obeys

$$\bar{j}^\mu = e\bar{\psi}\gamma^\mu\psi, \quad j^\mu{}_{;\mu} = \frac{1}{\sqrt{-g}}\partial_\mu\left(\sqrt{-g}\,\bar{j}^\mu\right) = f^{-4}\partial_\mu\left(f^4\bar{j}^\mu\right) = 0 . \tag{47}$$

(The symbol $\bar{j}$ is used for the invariant current for consistency with the notation in the discussion of EM.) This is consistent with $\bar{j}^\mu = f^{4\sigma}\tilde{j}^\mu$ and (46) only if $\sigma = -1$, which fixes the correct value of σ in (39):

$$\Gamma_\nu(x) = \frac{1}{2}\gamma_\nu \not{\partial}\phi . \tag{48}$$

6.3 A note on related work

Barut with others (Barut, A. O. & Duru, I. H., 1987; Barut, A. O. & Singh, L. P., 1995) give the spin connection equation (39) with $\sigma = 1$. Their equation is employed (Huang, 2005; Parashar, D., 1991) and otherwise widely cited elsewhere. An explicit equation of the form (39) with the correct sign is given has been given by Kovalyov (Kovalyov, M. & Légaré, M., 1989). Others (Birrell, N. D. & Davies, P. C. W., 1982; Villalba, V. M. & Percoco, U., 1989) bypass (39) and use the closed form result for the spin connection derived by Weinberg (Weinberg, S., 1972). That form can be shown to be compatible with (39) only for $\sigma = -1$. The result (42) has been given recently (Finster, F. & Reintjes, M., 2009), though the notation used by those authors is quite different.

6.4 Summary

It follows from the above that the Dirac equation in conformal spacetime with scale factor $f(x)$ is the same as the Dirac equation in Minkowski spacetime with the replacement

$$m \rightarrow f(x)m \tag{49}$$

whilst treating the current as conserved as usual (i.e. in Minkowski spacetime). It follows that an effective Lagrangian for the Dirac wavefunction in conformal spacetime is

$$I = -\int d^4x\, \bar{\psi}(x)\Big(\gamma^\alpha\big(i\hbar\partial_\alpha - eA_\alpha(x)\big) - f(x)m\Big)\psi(x) \tag{50}$$

where the γ^α are ordinary (Minkowski) gamma matrices. (Note that the scaling here differs from (Birrell & Davies, 1982); here the wavefunction is normalized for current conservation in Minkowski spacetime, un-weighted by the conformal factor.)

Though much has been written about the Dirac equation in conformal spacetime, this general result seems not to have been noticed, though there have been solutions given for the de Sitter case (see below) which effectively employ (50). The omission is probably due to the practice of working in the traditional Robertson-Walker coordinates rather than in conformal coordinates, which, especially in the case of curved space, tends to obscure the possibility of a reduction to (49).

7. Wavefunction in De Sitter spacetime

7.1 Feynman - Gell-Mann method of solution

Here we employ (45) to solve for the particular case of de Sitter spacetime with $f(x) = a(t) = 1/(Ht)$, regarded here as the asymptotic limit of the vacuum-dominated $K = 0$ Robertson-Walker cosmology. Given the above finding, henceforth we drop the tilde on the wavefunction and presume to be working solely in Minkowski spacetime with a dynamic mass. Then the free-space Dirac wavefunction (45) obeys

$$\big(i\gamma^\mu\partial_\mu - \lambda/t\big)\psi = 0; \quad \lambda = \frac{m_0 c^2}{\hbar H} \tag{51}$$

λ is a dimensionless number of order 10^{40}. We apply the Feynman - Gell-Mann method and make the substitution

$$\psi(t,\mathbf{x}) = \big(i\gamma^\mu\partial_\mu + \lambda/t\big)\phi(t,\mathbf{x}) . \tag{52}$$

Then ϕ satisfies

$$\big(i\gamma^\mu\partial_\mu - \lambda/t\big)\big(i\gamma^\mu\partial_\mu + \lambda/t\big)\phi = \big(\partial^2 + \lambda\big(\lambda + i\gamma^0\big)/t^2\big)\phi = 0 . \tag{53}$$

Separating out the spatial dependence of a single Fourier mode:

$$\phi(t,\mathbf{x}) = e^{i\mathbf{k}.\mathbf{x}}\phi(t;\mathbf{k}), \quad \psi(t,\mathbf{x}) = e^{i\mathbf{k}.\mathbf{x}}\psi(t;\mathbf{k}) \tag{54}$$

the bi-spinor $\phi(t;\mathbf{k}) = \chi(z); \quad z = |\mathbf{k}|t$ satisfies

$$\left(\frac{d^2}{dz^2}+1+\frac{\lambda\left(\lambda+i\gamma^0\right)}{z^2}\right)\chi(z)=0 \tag{55}$$

and now (52) gives

$$\psi(t,\mathbf{k})=\left(i\gamma^0\frac{\partial}{\partial z}-\hat{\mathbf{K}}+\frac{\lambda}{z}\right)\chi(z). \tag{56}$$

In the Dirac representation

$$\gamma^0=diag(1,1,-1,-1),\quad \chi^T=\left(u_+,\quad v_+,\quad u_-,\quad v_-\right) \tag{57}$$

where (u_+, v_+) is a positive energy spinor and (u_-, v_-) is a negative energy spinor, and

$$\psi(t,\mathbf{k})=\left(\begin{pmatrix}\lambda/z+i\partial_z & 0 \\ 0 & \lambda/z-i\partial_z\end{pmatrix}-\hat{\mathbf{K}}\right)\chi(z). \tag{58}$$

(In this notation u_+ and v_- are both spin up, u_- and v_+ are both spin down.) With (57), (55) becomes

$$\left(\frac{d^2}{dz^2}+1+\frac{\lambda\left(\lambda\pm i\right)}{z^2}\right)u_\pm=0. \tag{59}$$

The solution can be written in terms of Bessel functions which we choose to write as Hankel functions

$$u_\pm=\sqrt{z}H^{(1)}_{i\lambda\mp1/2}(z),\quad v_\pm=\sqrt{z}H^{(2)}_{i\lambda\mp1/2}(z) \tag{60}$$

Write the solutions of (56) as

$$\chi^T(z)=\sqrt{z}\left(H^{(1)}_{i\lambda-1/2}(z),\quad H^{(2)}_{i\lambda-1/2}(z),\quad H^{(1)}_{i\lambda+1/2}(z),\quad H^{(2)}_{i\lambda+1/2}(z)\right) \tag{61}$$

with implicit coefficients for the initial conditions. Using that (Abramowitz, M. & Stegun, I. A., 1965)

$$\frac{d}{dz}H^{(i)}_\nu(z)=\frac{\nu}{z}H^{(i)}_\nu(z)-H^{(i)}_{\nu+1}(z)$$
$$\Rightarrow\frac{d}{dz}\left(\sqrt{z}H^{(i)}_\nu(z)\right)=\frac{(\nu+1/2)}{\sqrt{z}}H^{(i)}_\nu(z)-\sqrt{z}H^{(i)}_{\nu+1}(z) \tag{62}$$
$$\Rightarrow(\lambda/z+i\partial_z)\left(\sqrt{z}H^{(i)}_{i\lambda-1/2}(z)\right)=-i\sqrt{z}H^{(i)}_{i\lambda+1/2}(z)$$

and that

$$\frac{d}{dz}H^{(i)}_\nu(z)=-\frac{\nu}{z}H^{(i)}_\nu(z)+H^{(i)}_{\nu-1}(z)$$
$$\Rightarrow\frac{d}{dz}\left(\sqrt{z}H^{(i)}_\nu(z)\right)=\frac{(-\nu+1/2)}{\sqrt{z}}H^{(i)}_\nu(z)+\sqrt{z}H^{(i)}_{\nu-1}(z) \tag{63}$$
$$\Rightarrow(\lambda/z-i\partial_z)\left(\sqrt{z}H^{(i)}_{i\lambda+1/2}(z)\right)=-i\sqrt{z}H^{(i)}_{i\lambda-1/2}(z)$$

then

$$
\begin{pmatrix} \lambda/z+i\partial_z & 0 \\ 0 & \lambda/z-i\partial_z \end{pmatrix}\chi(z) = \left(i\gamma^0\partial_z+\lambda/z\right)\chi(z) = -i\gamma^5\chi(z)\,. \tag{64}
$$

Putting this in (58) gives

$$
\psi(t,\mathbf{k}) = -\left(\hat{K}+i\gamma^5\right)\chi(z)\,. \tag{65}
$$

7.2 Independent solutions

We wish now to generalize (65) to give the four independent solutions. From (55) any linear combination of $\chi(z)$ that commutes with γ^0 will remain a solution of that equation. In order for the solution of the Dirac equation to retain the structure of (65) we would need also to have the linear combination commute with $\hat{K}$ and γ^5. These two constraints combined are that the linear combinations must commute with γ^a; $a \in \{0,1,2,3,5\}$. The possibilities include the identity and pairwise combinations of $\gamma^i\gamma^j$; i, $j \in \{1,2,3\}$. One possibility is the matrix Q

$$
Q := \begin{pmatrix} a & c & 0 & 0 \\ d & b & 0 & 0 \\ 0 & 0 & a & c \\ 0 & 0 & d & b \end{pmatrix} = a\left(1+i\gamma^1\gamma^2\right)/2+b\left(1-i\gamma^1\gamma^2\right)/2+c\left(i\gamma^2+\gamma^1\right)\gamma^3/2+d\left(i\gamma^2-\gamma^1\right)\gamma^3/2\,, \tag{66}
$$

hence

$$
\begin{aligned}
\psi(t,\mathbf{k}) &= \left(i\gamma^0\frac{\partial}{\partial z}-\hat{K}+\frac{\lambda}{z}\right)Q\chi(z) \\
&= Q\left(i\gamma^0\frac{\partial}{\partial z}-\hat{K}+\frac{\lambda}{z}\right)\chi(z) \\
&= -Q\left(\hat{K}+i\gamma^5\right)\chi(z)
\end{aligned} \tag{67}
$$

is a general solution, having 4 independent complex degrees of freedom. In order to relate the wavefunction to its initial state it will be more useful to express the constants as a vector:

$$
Q\chi(z) = \Lambda(z)q;\quad q^T = (a,b,c,d)\,. \tag{68}
$$

Re-arranging (66) one obtains

$$
\Lambda(z) = \sqrt{z}\begin{pmatrix} H^{(1)}_{i\lambda-1/2}(z) & 0 & H^{(2)}_{i\lambda-1/2}(z) & 0 \\ 0 & H^{(2)}_{i\lambda-1/2}(z) & 0 & H^{(1)}_{i\lambda-1/2}(z) \\ H^{(1)}_{i\lambda+1/2}(z) & 0 & H^{(2)}_{i\lambda+1/2}(z) & 0 \\ 0 & H^{(2)}_{i\lambda+1/2}(z) & 0 & H^{(1)}_{i\lambda+1/2}(z) \end{pmatrix}\,. \tag{69}
$$

With this, the solution (67) can be written

$$\psi(t,\mathbf{k}) = -\left(\hat{\mathbf{K}} + i\gamma^5\right)\Lambda(z)q . \tag{70}$$

7.3 Initial conditions

Let the initial conditions be that $\psi(t_0,\mathbf{k})$ is given. Then q can be found from

$$\psi(t_0,\mathbf{k}) = -\left(\hat{\mathbf{K}} + i\gamma^5\right)\Lambda(z_0)q . \tag{71}$$

Observing that

$$\left(\hat{\mathbf{K}} + i\gamma^5\right)^{-1} = -\frac{1}{2}\left(\hat{\mathbf{K}} + i\gamma^5\right) \tag{72}$$

then q is given by

$$q = \frac{1}{2}\Lambda^{-1}(z_0)\left(\hat{\mathbf{K}} + i\gamma^5\right)\psi(t_0,\mathbf{k}) . \tag{73}$$

Making use of the wronskian (Abramowitz & Stegun, 1965)

$$H_{\nu+1}^{(1)}(z)H_\nu^{(2)}(z) - H_\nu^{(1)}(z)H_{\nu+1}^{(2)}(z) = -\frac{4i}{\pi z} , \tag{74}$$

the inverse of (69) is found to be

$$\Lambda^{-1}(z) = \frac{i\pi}{4}\sqrt{z}\begin{pmatrix} -H_{i\lambda+1/2}^{(2)}(z) & 0 & H_{i\lambda-1/2}^{(2)}(z) & 0 \\ 0 & H_{i\lambda+1/2}^{(1)}(z) & 0 & -H_{i\lambda-1/2}^{(1)}(z) \\ H_{i\lambda+1/2}^{(1)}(z) & 0 & -H_{i\lambda-1/2}^{(1)}(z) & 0 \\ 0 & -H_{i\lambda+1/2}^{(2)}(z) & 0 & H_{i\lambda-1/2}^{(2)}(z) \end{pmatrix} . \tag{75}$$

With this, the wavefunction at arbitrary times can be expressed explicitly in terms of the initial state:

$$\psi(t,\mathbf{k}) = -\frac{1}{2}\left(\hat{\mathbf{K}} + i\gamma^5\right)\Lambda(z)\Lambda^{-1}(z_0)\left(\hat{\mathbf{K}} + i\gamma^5\right)\psi(t_0,\mathbf{k}) . \tag{76}$$

Cotaescu and Crucean (Coaescu, I. I. & Crusean, C., 2008) claimed to be the first to give a solution of (51) involving Hankel functions of complex order.

8. Effects of the conformal singularity

8.1 Asymptotic behavior far from the singularity

The region far from the singularity is characterized by $z \gg 1$ where $z = |\mathbf{k}|t$. Given the conformal time to the boundary computed in (19) is $t_{cb} = 47.2$ Gyr, for a particle to be in this region at the present time requires its speed $|\mathbf{v}|$ satisfy

$$|\mathbf{k}|ct_{cb} \gg 1 \Rightarrow |\mathbf{v}| \gg \frac{\hbar}{mct_{cb}} = \frac{c}{\omega_c t_{cb}} . \tag{77}$$

In practice in means the speed with respect to the Cosmological Frame must satisfy $|\mathbf{v}| \gg 10^{-31}$ m/s. For all practical purposes therefore, all matter in the present era is in the 'far field' of the future boundary.

We suppose that the initial condition is given a long way from the singularity, i.e. where $|z_0| \gg 1$. This is not the usual way of doing things, but here the conformal boundary is in the future and the wavefunction there will be presumed determined by propagation from an initially known state much earlier. Actually, the asymptotic expansions for the Hankel functions are given for large *positive* argument, which in our case is far in the future on the other side of the conformal boundary. Rather than try to change things around, we will compute the behavior of the wavefunction as it approaches the boundary from above – i.e. going backwards in time – assuming that the initial conditions (q) is given further into the future, with the asymptotic behavior at the boundary to be determined. In that case the mass is negative and λ in (51) is large and negative (and real).

The magnitude of the Hankel functions are independent of order in the limit of large magnitude, so all the terms in (75) have equal weight. Use (Abramowitz & Stegun, 1965)

$$\sqrt{z_0}H_\alpha^{(1)}(z_0) \to \sqrt{\frac{2}{\pi}}e^{i(z_0-\alpha\pi/2-\pi/4)}, \quad \sqrt{z_0}H_\alpha^{(2)}(z_0) \to \sqrt{\frac{2}{\pi}}e^{-i(z_0-\alpha\pi/2-\pi/4)}. \tag{78}$$

Specifically:

$$\sqrt{z_0}H_{i\lambda+1/2}^{(1)}(z_0) \to -i\sqrt{\frac{2}{\pi}}e^{\lambda\pi/2+iz_0}, \quad \sqrt{z_0}H_{i\lambda+1/2}^{(2)}(z_0) \to i\sqrt{\frac{2}{\pi}}e^{-\lambda\pi/2-iz_0}$$

$$\sqrt{z_0}H_{i\lambda-1/2}^{(1)}(z_0) \to \sqrt{\frac{2}{\pi}}e^{\lambda\pi/2+iz_0}, \quad \sqrt{z_0}H_{i\lambda-1/2}^{(2)}(z_0) \to \sqrt{\frac{2}{\pi}}e^{-\lambda\pi/2-iz_0} \tag{79}$$

so $\Lambda^{-1}(z_0)$ in tends to

$$\Lambda^{-1}(z_0) \to i\sqrt{\frac{\pi}{8}}\begin{pmatrix} e^{-\lambda\pi/2-iz_0} & 0 & ie^{-\lambda\pi/2-iz_0} & 0 \\ 0 & e^{\lambda\pi/2+iz_0} & 0 & -ie^{\lambda\pi/2+iz_0} \\ e^{\lambda\pi/2+iz_0} & 0 & -ie^{\lambda\pi/2+iz_0} & 0 \\ 0 & e^{-\lambda\pi/2-iz_0} & 0 & ie^{-\lambda\pi/2-iz_0} \end{pmatrix}. \tag{80}$$

8.2 Wavefunction near the conformal singularity

Near the boundary z is small (and positive here) whereupon the Hankel functions approach the limiting forms (Abramowitz & Stegun, 1965)

$$\sqrt{z}H_\alpha^{(1)}(z) \to -\frac{i}{\pi}\left(\frac{2}{z}\right)^{\alpha-1/2}\Gamma(\alpha), \quad \sqrt{z}H_\alpha^{(2)}(z) \to \frac{i}{\pi}\left(\frac{2}{z}\right)^{\alpha-1/2}\Gamma(\alpha); \quad \mathrm{Re}(\alpha) > 0. \tag{81}$$

Specifically:

$$\sqrt{z}H_{i\lambda+1/2}^{(1)}(z) \to -\frac{i}{\pi}\left(\frac{2}{z}\right)^{i\lambda}\Gamma(i\lambda+1/2), \quad \sqrt{z}H_{i\lambda+1/2}^{(2)}(z) \to \frac{i}{\pi}\left(\frac{2}{z}\right)^{i\lambda}\Gamma(i\lambda+1/2). \tag{82}$$

But (81) is valid only for real α, so we will need to express the $H^{(j)}_{i\lambda-1/2}(z)$ differently before applying those limits. Using that

$$H^{(j)}_{i\lambda-1/2}(z) = -H^{(j)}_{i\lambda+3/2}(z) + \frac{2i\lambda+1}{z} H^{(j)}_{i\lambda+1/2}(z) \tag{83}$$

then

$$\sqrt{z} H^{(1)}_{i\lambda-1/2}(z) = -\sqrt{z} H^{(1)}_{i\lambda+3/2}(z) + \frac{2i\lambda+1}{\sqrt{z}} H^{(1)}_{i\lambda+1/2}(z)$$
$$\to \frac{i}{\pi}\left(\frac{2}{z}\right)^{i\lambda+1}\Gamma(i\lambda+3/2) - \frac{2i\lambda+1}{z}\frac{i}{\pi}\left(\frac{2}{z}\right)^{i\lambda}\Gamma(i\lambda+1/2). \tag{84}$$
$$= \frac{i}{\pi}\left(\frac{2}{z}\right)^{i\lambda+1}\Gamma(i\lambda+1/2)$$

Similarly

$$\sqrt{z} H^{(2)}_{i\lambda-1/2}(z) \to -\frac{i}{\pi}\left(\frac{2}{z}\right)^{i\lambda+1}\Gamma(i\lambda+1/2). \tag{85}$$

Putting (85), (84) and (82) into (69) gives

$$\Lambda(z) \to \frac{i}{\pi}\left(\frac{2}{z}\right)^{i\lambda}\Gamma(i\lambda+1/2)\begin{pmatrix} 2/z & 0 & -2/z & 0 \\ 0 & -2/z & 0 & 2/z \\ -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}. \tag{86}$$

Near the boundary therefore

$$\Lambda(z) \to \frac{i}{\pi}\left(\frac{2}{z}\right)^{i\lambda+1}\Gamma(i\lambda+1/2)\begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{87}$$

Since the entries in this matrix now have equal weight with respect to λ, the entries in the matrix $\Lambda^{-1}(z_0)$ can be assessed accordingly. The region $z > 0$ (to which the expansions are restricted) corresponds to the post-singularity half-space, in which the rest mass is negative. Therefore $\lambda < 0$, $|\lambda| \gg 1$ and only the exponentials of the form $\exp(-\lambda\pi/2)$ survive. The inverse $\Lambda^{-1}(z_0)$ then simplifies to

$$\Lambda^{-1}(z_0) \to i\sqrt{\frac{\pi}{8}}e^{|\lambda|\pi/2}e^{-iz_0}\begin{pmatrix} 1 & 0 & i & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & i \end{pmatrix} \tag{88}$$

where we have used that $\exp(-\lambda\pi/2) = \exp(|\lambda|\pi/2)$ to make the large size of the factor explicit. Combining (88) and (87):

$$\Lambda(z)\Lambda^{-1}(z_0) \rightarrow -\frac{1}{z\sqrt{2\pi}}\left(\frac{2}{z}\right)^{i\lambda} e^{|\lambda|\pi/2} e^{-iz_0} \Gamma(i\lambda+1/2)\begin{pmatrix} 1 & 0 & i & 0 \\ 0 & 1 & 0 & i \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \qquad (89)$$

The gamma function has asymptotic behavior (Abramowitz & Stegun, 1965)

$$\left|\Gamma(i\lambda+x)\right| \rightarrow \sqrt{2\pi}\,|y|^{x-1/2}\,e^{-\pi|\lambda|/2} \Rightarrow \left|\Gamma(i\lambda+1/2)\right| \rightarrow \sqrt{2\pi}e^{-\pi|\lambda|/2}\,, \qquad (90)$$

so (89) can be written

$$\Lambda(z)\Lambda^{-1}(z_0) \rightarrow \frac{1}{z}e^{i\phi(z,z_0,\lambda)}\begin{pmatrix} 1 & 0 & i & 0 \\ 0 & 1 & 0 & i \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \frac{1}{2z}e^{i\phi(z,z_0,\lambda)}\left(1+\gamma^0\right)\left(1+i\gamma^5\right) \qquad (91)$$

where ϕ is a (real) phase. Putting this into (76), the wavefunction tends to a limit which, in terms of its initial value far from the boundary, is

$$\psi(t,\mathbf{k}) \rightarrow -\frac{1}{4z}e^{i\phi(z,z_0,\lambda)}\left(\hat{\mathbf{K}}+i\gamma^5\right)\left(1+\gamma^0\right)\left(1+i\gamma^5\right)\left(\hat{\mathbf{K}}+i\gamma^5\right)\psi(t_0,\mathbf{k})\,. \qquad (92)$$

Defining the projection $P(\mathbf{k})$

$$P(\mathbf{k}) = \frac{1}{2}\left(1-\gamma^0\right)\left(1+\hat{\mathbf{K}}\right) = \frac{1}{k}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -k_z & -k_x+ik_y & k & 0 \\ -k_x-ik_y & k_z & 0 & k \end{pmatrix} \qquad (93)$$

this becomes

$$\psi(t,\mathbf{k}) \rightarrow \frac{1}{z}e^{i\phi(z,z_0,\lambda)}P(\mathbf{k})\psi(t_0,\mathbf{k})\,. \qquad (94)$$

8.3 On the reduction to two Dirac components
We see from (93) that although all four components may be defined (as non-zero) away from the boundary, they are rolled up into just two components by the time they reach the boundary; only the *negative* energy components of the Dirac wavefunction are non-zero there.

With appropriately chosen asymptotic behavior for the Hankel functions the same procedure can be applied to the pre-boundary wavefunction, wherein the rest mass and therefore λ are positive. In that case one finds that only *positive* energy states arrive at the boundary - whatever the values of the 4 components specified in the 'initial conditions' away from the boundary.

On 'our' side at $t = 0_-$ the negative energy components are zero at the singularity. On the other side at $t = 0_+$ the positive energy components are zero. As it approaches the singularity the phase oscillates rapidly, infinitely quickly so at the boundary. Moving away from the singularity towards the present era – i.e. backwards in time – the appearance of non-zero negative energy components is due to subsequent interactions in which the initially pure positive energy state becomes mixed, and all four components in the Dirac wavefunction bi-spinor become occupied.

Note that this behavior by itself is *not* a boundary condition on the wavefunction. The negative energy states decay to zero as a consequence of intrinsic properties of the Dirac equation. If a boundary condition *were* to be imposed, it could only be on the positive energy states at $t = 0_-$ - regarded as a future boundary condition, and on negative energy states at $t = 0_+$ - regarded as an historical boundary condition for development of the post-singularity universe.

The analysis above culminating in (92) gives that only the positive energy components of the wavefunction components are non-zero at $t = 0_-$. It is expected that this behavior carries over into the second quantized theory so that electrons and positrons arrive at the boundary with zero velocity relative to the Hubble frame.

8.4 Hubble drag

The 4-current in the Dirac representation is

$$\left\{ j^\mu \right\} = \psi^\dagger \gamma^0 \left\{ \gamma^\mu \right\} \psi = \psi^\dagger \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \left(\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ -\boldsymbol{\sigma} & 0 \end{pmatrix} \right) \psi = \psi^\dagger \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ \boldsymbol{\sigma} & 0 \end{pmatrix} \right) \psi . \qquad (95)$$

As a result of the loss of the two negative energy components the 3-current vanishes:

$$\rho = e\left(\left| u_+ \right|^2 + \left| v_+ \right|^2 \right), \quad \mathbf{j} = \mathbf{0} . \qquad (96)$$

An effect of the expansion is a 'Hubble Drag' that acts to bring all Dirac matter to rest in the Hubble frame. The same outcome is predicted of classical matter by the action

$$I = -m \int dt \, a(t) \sqrt{1 - \mathbf{v}^2(t)}; \quad \mathbf{v}(t) := d\mathbf{x}(t) / dt . \qquad (97)$$

Here t is the conformal time and $a \sim 1/t$. The effect of Hubble drag on the current is shown in Figure 2. A fuller discussion of this compared with the more traditional classical action

$$I = -m \int \sqrt{dx^\mu dx^\nu g_{\mu\nu}(x)} = -m \int dt \left| a(t) \right| \sqrt{1 - \mathbf{v}^2(t)} \qquad (98)$$

can be found in (Ibison, 2010).

9. Future cosmological boundary condition

9.1 Relation between pre and post singularity universes

We will assume the laws of physics do not change either side of the singularity and in particular that Maxwell's equations and the single particle Dirac equation are obeyed everywhere, including at and through the conformal singularity. An assessment of the
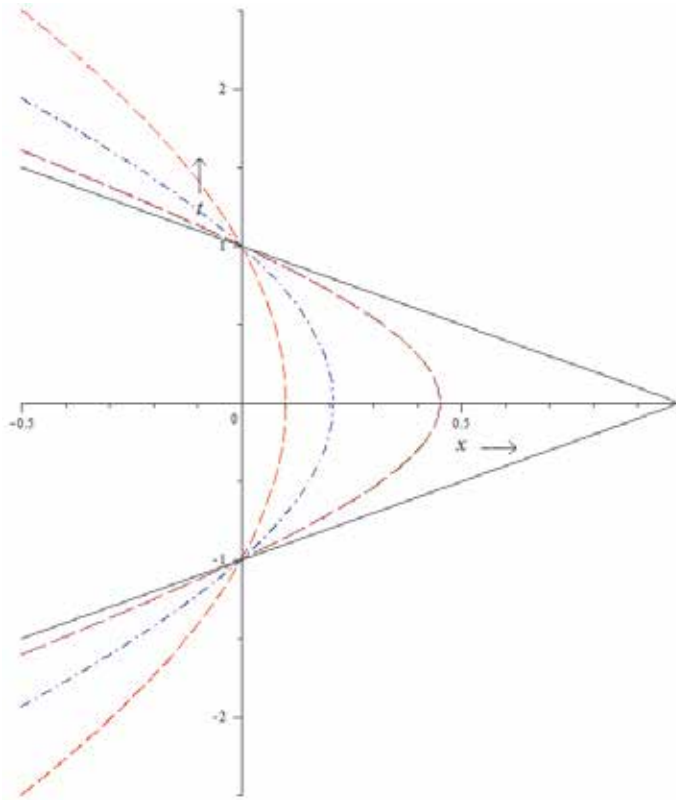
Fig. 2. Geodesics of matter in the de Sitter limit of evolution for varying initial speeds specified far from the Conformal Singularity, which is at $t = 0$. The solid line is the geodesic for light speed as a limit approached from below. The time axis measures conformal time.

relationship between the pre and post singularity universes should take account of the following two important facts:

i.   The scale factor is anti-symmetric across the boundary, and

ii.   The pre-singularity wavefunction possesses symmetric partners across the boundary.

Because the Friedmann equation is symmetric about the boundary, the post-singularity evolution is a time-reversed version of the expansion on our side. It is hard to see how minor deviations can be tolerated at the particle level without damaging the symmetry of the scale factor, even though the latter signifies the development of only the most course-grain level of the Cosmological fluid. There is no scope in the Friedmann equation for minor deviations from perfect symmetry, so it is to be expected that this state of affairs be upheld consistently at every level, down to the individual particle wavefunctions. From consideration of the Friedmann equation alone therefore, one expects to find that the pre and post singularity Dirac wavefunctions are symmetric partners such that their contributions to the stress-energy at fixed **x** from both t and $-t$ (relative to the singularity at $t = 0$) are identical.

We wish to make clear that there is no imperative from solving the Dirac equation alone that the pre and post singularity wavefunctions have any relationship. Indeed, the opposite seems true at first; the wavefunctions on either side of the singularity have no non-zero

components in common and so are completely decoupled. Thus the input from the Friedmann equation that the two universes be related amounts to an extraneous constraint on the relationship between the two wavefunctions. It means that the two components that 'survive' on each side must somehow be related.

9.2 The 'No-Copy' assumption

In the following we seek a picture that is consistent with both the symmetries of the Dirac equation in conformal spacetime and the Friedmann equation. A reasonable conclusion is that the post singularity universe is in some sense a copy of the pre-singularity universe. The most efficient explanation from an Ockham's Razor point of view is that there is no copy: the post-singularity universe and the post-singularity universe are physically the same universe. Then the geodesics mirrored in the conformal singularity in Figure 2 are to be taken as standing for the development of the ensemble as a whole, presumed therefore to be faithfully mirrored in detail at every level.

If we put the conformal singularity at $t = 0$, then the no copy constraint says that the universe at t and $-t$, are the same, and therefore the state of affairs described physically on either side are related by an unobservable transformation. It does not have to mean that Cosmology is cyclic when parsed with monotonically increasing t – it would only be so if the same condition were applied at the Big Bang. Given this, below we enumerate the possibilities consistent with the no-copy assumption.

9.3 Time-reversed image

The post-singularity wavefunction will be an identical copy of the pre-singularity wavefunction if the coordinate system is folded back upon itself in such a manner that the pre and post singularity coordinates $(t,\mathbf{x})$ and $(-t,\mathbf{x})$ are physically the same point. Then, given a pre-singularity solution $\psi(-t,\mathbf{x})$, $t > 0$ the 'no-copy' requirement is that there must exist a post-singularity solution $\psi(t,\mathbf{x}) = \eta\psi(-t,\mathbf{x})$, $t > 0$ where η is a constant unobservable phase. We would prefer these be compatible with the point of view that the wavefunction crosses the boundary propagating according to business as usual from the perspective of the Dirac equation (51), and then folding the space along the conformal singularity to find the post-singularity evolution is identical with the pre-singularity evolution. There is a problem however: From the right-most column of Table 1 one sees that there is no symmetry corresponding to $\psi(t,\mathbf{x}) = \eta\psi(-t,\mathbf{x})$, $t > 0$ because time reversal introduces γ^0 which can be removed only with a parity inversion. Hence this possibility can be discounted.

9.4 Parity inversion

Motivated by CPT invariance of QED, we seek a 'no-copy' interpretation of the post-singularity universe in which $(t,\mathbf{x})$ and $(-t,-\mathbf{x})$ are physically the same points. The 'no copy' requirement then becomes $\psi(t,\mathbf{x}) = \eta\psi(-t,-\mathbf{x})$, $t > 0$. Reading off from Table 1 the possibilities are

$$\eta = \mathrm{P}_\rho\mathrm{T}_\rho\mathrm{M}_- \text{ or } \mathrm{P}_\rho\mathrm{T}_\rho\mathrm{C}_+ \text{ or } \mathrm{P}_\rho\mathrm{T}_{\bar\rho}\mathrm{M}_+ \text{ or } \mathrm{P}_\rho\mathrm{T}_{\bar\rho}\mathrm{C}_-; \quad \rho \in \left[+,-\right] \quad ,\bar\rho = -\rho \ . \tag{99}$$

The identical universe requires also that

$$A^\mu(-x) = A^\mu(x) \tag{100}$$

with no room for absorption of an arbitrary factor. Note this is a point symmetry, not a local one. With reference to Table 1, charge inversion is therefore discounted, leaving only

$$\eta = P_\rho T_\rho M_- \text{ or } P_\rho T_{\bar\rho} M_+ \quad . \tag{101}$$

This is as far as we can proceed unambiguously. Possibly there is no physically meaningful distinction to be made between the alternatives, in which case attempting to isolate one is a waste of effort. Acknowledging this risk, we give the following argument in favor of a particular combination. We would like to ascribe the process of mass inversion solely with propagation of the Dirac function across the boundary, in accord with the effect of the scale factor in the Dirac equation. As we have said, the singularity swaps the positive and negative energy solutions, which is achieved by both M_+ and M_- (proportional to γ^5 and γ^2 respectively). Despite the singular behavior of the Dirac wavefunction at the boundary we would prefer to regard this exchange as the limit of an analytical process, which does not seem possible if the operation is anti-linear. This singles out M_+. Similar reasoning concerning the coordinate folding at the singularity singles out the real operations $P_+ T_-$. This reasoning isolates $\eta = P_+ T_- M_+$ as the only viable candidate. There is a close connection with CPT invariance here. Unlike the latter, we have set ourselves the constraint that the wavefunction and the EM fields transform back to themselves with no room in the latter for a sign inversion which is the reason mass inversion was favored over charge conjugation.

As briefly discussed in an earlier work (Ibison, 2010) a parity inversion can be achieved in the de Sitter spacetime at the time of a pair of conformal singularities by identification of the two hypersurfaces of co-dimension 1 there (Calabi, E. & Marcus, L., 1962; Hawking, S. W. & Ellis, G. F. R., 1973). In our case, the parity inversion and the 'no-copy' assumption can be achieved through the identification of two 4-volumes. In the case of the first of (3) and putting the conformal singularity at $t = 0$, the two volumes are inversions of each other through the origin. If there are no topological implications of the Big Bang then the topology is

$$\mathbb{R}^4 / \sim, \quad x \sim -x \; \forall x \in \mathbb{R}^4. \tag{102}$$

In practice the conformal time since the Big Bang is finite so this manifold is only partially occupied by our universe. A wavefunction in our half-space $t < 0$ is unaffected by this identification, though at $t = 0$ it requires $\psi(0,\mathbf{x}) = \psi(0,-\mathbf{x})$ for all $\mathbf{x}$. Not wanting to favor any particular location, this implies $\psi(0,\mathbf{x}) = $ constant - a future boundary condition on the development of the Dirac wavefunction in our half-space. In the second quantized theory it is expected that this condition will take a form that requires every particle be annihilated by its anti-particle at the boundary.

The effects of the boundary condition will not generally be noticeable far from boundary, and similar to that of EM fields as discussed in more detail elsewhere (Ibison, 2010). Since only Fourier modes having wavelengths approaching that of the distance to the boundary are appreciably affected, effects become noticeable locally only at very low speeds and perhaps also at very low accelerations. The possibility of a connection with Modified Newtonian Dynamics (MOND) should be mentioned because the threshold acceleration for the onset of MOND effects is around $a_0 \sim 10^{-10} \, \text{m/s}^2 \sim cH$ as most recently reported (Swaters, R. A., Sanders, R. H., and McGaugh, S. S., 2010) and H sets the length scale (specifically c/H) at which wavelengths one expects to notice departures from standard

theories that do not include a future boundary. Discussion of the implications for entropy and the second law of thermodynamics are omitted here, though it is hard to see how these could escape unmodified.

10. Summary

We examined in detail the effect of the conformal factor on the discrete symmetries usually present in the Dirac theory and analyzed the behavior of the Dirac wavefunction near the future conformal singularity. We found that only two of the 4 bi-spinor components survive to the boundary, one effect of which is to cause the geodesics to experience a drag towards the Hubble frame. We then presented arguments for the existence of a genuine boundary condition on the wavefunction at the future conformal singularity based upon the assumption there does not exist a redundant copy of the universe in the post-conformal singularity era. The future boundary condition affects the spectral decomposition of the wavefunction and this might be locally testable, i.e. at the present time.

11. Appendix A: discrete symmetries in Minkowski spacetime

Equation Section (Next). In order to catalog the effects of the conformal factor let us first review the discrete symmetries as they would be in Minkowski spacetime, i.e. with $f(x) = 1$ in (1).

$$\left(\gamma^{\mu}\left(-i\partial_{\mu} + eA_{\mu}(x)\right) + m\right)\psi(x) = 0 . \tag{A1}$$

We deal first with Lorentz scalars and vectors, and subsequently with the Dirac equation. This appendix is a review and compilation of material which can be found, for example, in (Itzykson, C. & Zuber, J.-B., 1985; Peskin, M. E. & Schroeder, D. V., 1995; Weinberg, S., 2005).

A.1 Lorentz quantities

Let C,P,T,M denote charge parity time and mass inversions. By mass here we mean rest-mass, which is a Lorentz scalar. Mass and charge inversions are internal to the particle and not part of the Minkowski geometry and not, therefore, members of the Lorentz Group. Let I be any member of this set of discrete inversions; $I \in [P,T,C,M]$, $I^2 = 1$. Fundamentally their effects are just to change signs of the associated quantities as summarized in Table A1.

	e	m	x
P	+	+	$\tilde{x}$
T	+	+	$-\tilde{x}$
C	-	+	x
M	+	-	x

Table A1. Effects of charge, parity, time and mass inversions on e, m, and x.

Here, '–' denotes a sign inversion and

$$\tilde{x} = \left\{ h^\mu{}_\nu x^\nu \right\}, \quad h^\mu{}_\nu := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{A2}$$

The Lorentz scalar fields transform straightforwardly as

$$\text{\scriptsize I}\left[f(x;e,m) \right] = f\left(\text{\scriptsize I}^{-1}[x]; \text{\scriptsize I}^{-1}[e], \text{\scriptsize I}^{-1}[m] \right) = f\left(\text{\scriptsize I}[x]; \text{\scriptsize I}[e], \text{\scriptsize I}[m] \right), \tag{A3}$$

the second step following because $\text{\scriptsize I}^2 = 1$. To determine the effect on a Lorentz vector we could perhaps take as our template the derivative of a scalar field associated perhaps with a charge e and mass m: $f(x;e,m)$. It's transformation is easily determined from the substitutions $\mathbf{x} \to \mathbf{x}' = -\mathbf{x}$ so $\nabla \to -\nabla$ and $t \to t' = -t$ so $\partial / \partial t \to -\partial / \partial t$ [3]:

$$\begin{aligned} \text{\scriptsize M}\left[\partial_\mu f(x;e,m) \right] &= \partial_\mu f(x;e,-m) \\ \text{\scriptsize C}\left[\partial_\mu f(x;e,m) \right] &= \partial_\mu f(x;-e,m) \\ \text{\scriptsize P}\left[\partial_\mu f(x;e,m) \right] &= h_\mu{}^\nu \partial_\nu f(\tilde{x};e,m) \\ \text{\scriptsize T}\left[\partial_\mu f(x;e,m) \right] &= -h_\mu{}^\nu \partial_\nu f(-\tilde{x};e,m) \end{aligned} \tag{A4}$$

However, it is traditional to regard the zeroth component of a Lorentz vector as something that keeps its sign even under time reversals. For example the retarded Liénard-Wiechert potentials of a source following a path $\mathbf{x}_s(t)$ are sometimes written

$$A^\mu(t,\mathbf{x}) = \frac{e\left(1, d\mathbf{x}_s(t')/dt'\right)}{\left|\mathbf{x} - \mathbf{x}_s(t')\right| - \left(\mathbf{x} - \mathbf{x}_s(t')\right).d\mathbf{x}_s(t')/dt'}; \quad t - t' = \left|\mathbf{x} - \mathbf{x}_s(t')\right| \tag{A5}$$

which does not have the transformation property $\text{\scriptsize T}\left[A_\mu(x) \right] = -h_\mu{}^\nu A_\nu(-\tilde{x})$ that would be inferred from (A4). This is easiest to see in the limit of a static charge:

$$\phi(t,\mathbf{x}) = \frac{e}{\left|\mathbf{x} - \mathbf{x}_s\right|} \tag{A6}$$

---

[3] Another way of thinking about this is to regard the parity operation for example as converting the contra-variant to the covariant form and vice-versa:

$$\text{\scriptsize P}\left[V^\mu(x) \right] = V_\mu(\tilde{x}) \text{ and } \text{\scriptsize P}^2 = 1 \Rightarrow \text{\scriptsize P}\left(V_\mu(x) \right) = V^\mu(\tilde{x}).$$

However this has the effect of scrambling the Lorentz indexes in a product of a vector with a gamma matrix (because the latter are treated as universal constants):

$$\text{\scriptsize P}\left[\gamma^\mu V_\mu(x) \right] = \gamma^\mu \text{\scriptsize P}\left[V_\mu(x) \right] = \gamma^\mu V^\mu(\tilde{x}).$$

Here we prefer to treat the transformation matrix as a mixed tensor as defined in (A2).

which shows no capacity for changing sign under time reversal. The effects of time-reversal are felt instead in the space part, due to the presence there of derivatives with respect to time. Bearing this in mind, and taking into account also that the dependence on charge is known explicitly: $A_\mu(x) = A_\mu(x;e) = \text{sgn}(e)A_\mu(x;|e|)$, the inversions of the vector potential are

$$
\begin{aligned}
&\text{M}\left[A_\mu(x)\right] = A_\mu(x) \\
&\text{C}\left[A_\mu(x)\right] = -A_\mu(x) \\
&\text{P}\left[A_\mu(x)\right] = h_\mu{}^\nu A_\nu(\tilde{x}) \\
&\text{T}\left[A_\mu(x)\right] = h_\mu{}^\nu A_\nu(-\tilde{x})
\end{aligned}
\qquad (A7)
$$

We will see below that $\text{T}\left[A_\mu(x)\right] = h_\mu{}^\nu A_\nu(-\tilde{x})$ is consistent with the standard interpretation of the time reversal operator on the Dirac wavefunction and, consequently, on the 4-current.

A.2 Gamma matrices and slashed vectors

In discussing whether or not these symmetries exist the gamma matrices are treated here as universal constants, and not subject to or affected by the inversions. It will be useful in the following to have in hand the inversion of the slashed vectors:

$$
\begin{aligned}
&\text{M}\left[\gamma^\mu A_\mu(x)\right] = \gamma^\mu A_\mu(x) \\
&\text{C}\left[\gamma^\mu A_\mu(x)\right] = -\gamma^\mu A_\mu(x) \\
&\text{P}\left[\gamma^\mu A_\mu(x)\right] = \gamma^\mu h_\mu{}^\nu A_\nu(\tilde{x}) = \gamma^{\mu\dagger} A_\mu(\tilde{x}) = \gamma^0 \gamma^\mu \gamma^0 A_\mu(\tilde{x}) \\
&\text{T}\left[\gamma^\mu A_\mu(x)\right] = \gamma^0 \gamma^\mu \gamma^0 A_\mu(-\tilde{x})
\end{aligned}
\qquad (A8)
$$

and

$$
\begin{aligned}
&\text{M}\left[\gamma^\mu \partial_\mu\right] = \gamma^\mu \partial_\mu \\
&\text{C}\left[\gamma^\mu \partial_\mu\right] = \gamma^\mu \partial_\mu \\
&\text{P}\left[\gamma^\mu \partial_\mu\right] = \gamma^0 \gamma^\mu \gamma^0 \partial_\mu \\
&\text{T}\left[\gamma^\mu \partial_\mu\right] = -\gamma^0 \gamma^\mu \gamma^0 \partial_\mu
\end{aligned}
\qquad (A9)
$$

Inversion operations are symmetries of the Dirac equation if there exists an $\text{I}\left[\psi(x)\right]$ compliant with (A2) for which

$$
\text{I}\left[\left(\gamma^\mu\left(-i\partial_\mu + eA_\mu(x)\right) + m\right)\psi(x)\right] = \text{I}\left[\gamma^\mu\left(-i\partial_\mu + eA_\mu(x)\right) + m\right]\text{I}\left[\psi(x)\right] = 0 \qquad (A10)
$$

where $\text{I}\left[\psi(x)\right]$ can be written in terms of $\psi(x)$.

A.3 Linear and anti-linear inversion operations on the Dirac equation

Two possibilities are allowed for: I is linear and unitary, or anti-linear and anti-unitary. Let us distinguish between these two with subscripts as follows

$$
\begin{aligned}
I_+[i] &= i, \quad I_+\big[\psi(x)\big] = U_+\psi\big(I_+^{-1}[x]\big) = U_+\psi\big(I_+[x]\big) \\
I_-[i] &= -i, \quad I_-\big[\psi(x)\big] = U_-\psi^*\big(I_-^{-1}[x]\big) = U_-\psi^*\big(I_-[x]\big)
\end{aligned}
\tag{A11}
$$

where U_+, U_- are 4x4 matrices. The coordinates are insensitive to the distinction:

$$
I_+[x] = I_-[x] = I[x].
\tag{A12}
$$

It will be useful to write U_- in terms of U_+ so that just one of the two cases need be solved for. Applying (A11) and (A12) to (A10) gives

$$
\Big(\gamma^\mu{}_{I_+}\big[-i\partial_\mu + eA_\mu(x)\big] + I[m]\Big)U_+\psi\big(I[x]\big) = 0
\tag{A13}
$$

and

$$
\Big(\gamma^\mu{}_{I_-}\big[-i\partial_\mu + eA_\mu(x)\big] + I[m]\Big)U_-\psi^*\big(I[x]\big) = \Big(\gamma^\mu{}_{I_+}\big[i\partial_\mu + eA_\mu(x)\big] + I[m]\Big)U_-\psi^*\big(I[x]\big) = 0.
\tag{A14}
$$

Taking the conjugate of the latter:

$$
\Big(\gamma^{\mu*}{}_{I_+}\big[-i\partial_\mu + eA_\mu(x)\big] + I[m]\Big)U_-^*\psi\big(I[x]\big) = 0.
\tag{A15}
$$

We introduce a charge conjugation matrix C defined by the property

$$
-\gamma^{\mu T} = C\gamma^\mu C^{-1}
\tag{A16}
$$

C depends on the representation of the γ^μ. We will need

$$
\gamma^{\mu*} = -\gamma^0 C\gamma^\mu C^{-1}\gamma^0 = \gamma^0 C\gamma^5\gamma^\mu\gamma^5 C^{-1}\gamma^0
\tag{A17}
$$

with which (A15) can be written

$$
\Big(\gamma^\mu{}_{I_+}\big[-i\partial_\mu + eA_\mu(x)\big] + I[m]\Big)\gamma^5 C^{-1}\gamma^0 U_-^*\psi\big(I[x]\big) = 0.
\tag{A18}
$$

Comparing with (A13) one infers that

$$
U_+\psi\big(I[x]\big) = \kappa\gamma^5 C^{-1}\gamma^0 U_-^*\psi\big(I[x]\big) \Rightarrow U_-^* = \eta^*\gamma^0 C\gamma^5 U_+ \Rightarrow U_- = \eta\gamma^{0*}C^*\gamma^{5*}U_+^*
\tag{A19}
$$

where η and κ are phase factors. In the Dirac representation

$$
C = i\gamma^2\gamma^0 = C^* \Rightarrow C^{-1} = i\gamma^0\gamma^2 = -C
\tag{A20}
$$

and so (A19) is

$$
U_- = \eta\gamma^0 C\gamma^5 U_+^* = i\eta\gamma^0\gamma^2\gamma^0\gamma^5 U_+ = \lambda\gamma^2\gamma^5 U_+^*
\tag{A21}
$$

where λ is another phase factor. In the following we look at each inversion operation in detail, summarizing the results in Table A2.

A.4 Parity inversion
From (A8) and (A9),

$$\text{P}_+ \left[\gamma^\mu \left(-i\partial_\mu + eA_\mu(x) \right) + m \right] = \gamma^0 \gamma^\mu \gamma^0 \left(-i\partial_\mu + eA_\mu(\tilde{x}) \right) + m . \tag{A22}$$

We see that unless $A_\mu(\tilde{x}) = A_\mu(x)$, P_+ is not a symmetry of the single particle Dirac equation. (An example where it is a symmetry is a central potential, for which $A_\mu(x) = \left(e / |\mathbf{x}|, \mathbf{0} \right) = A_\mu(\tilde{x})$). Hence the dynamics are not generally invariant under parity inversion, though the kinematics of the free particle may be. Note we are not investigating here whether or not parity inversion is a symmetry which, if applied to all of physics, including therefore the sources of A_μ, leaves the dynamics unchanged. That is, we are not looking here for a *systematic symmetry* in the sense defined earlier. Instead, we are asking if there is a symmetric partner for a solution in a given field, which field is held constant whilst asking the question.

On the other hand, (A22) tells us that linear parity inversion is a *local symmetry* of the free Dirac particle equation with

$$\text{P}_+ \left[\psi(x) \right] = \eta \gamma^0 \psi(\tilde{x}) . \tag{A23}$$

That is, for every solution $\psi(x)$ there exists another solution $\eta \gamma^0 \psi(\tilde{x})$. Here and in the following η stands for a continuously re-definable phase factor which in each case we will choose so that the operator squares to 1. Hence in this case $\eta = 1$. We can read off from (A19) that the anti-linear version of parity inversion is

$$\text{P}_- \left[\psi(x) \right] = \eta \left(\gamma^0 C \gamma^5 \gamma^0 \psi(\tilde{x}) \right)^* . \tag{A24}$$

Using (A21) with (A23), in the Dirac representation this is

$$\text{P}_- \left[\psi(x) \right] = \gamma^2 \gamma^5 \gamma^0 \psi^*(\tilde{x}) \tag{A25}$$

(with η chosen so $\text{P}_-^{\ 2} = 1$). We now examine the relationship between the energy of the original particle and its parity inverted partner. A stationary particle solving (2) varies in time as $\exp(-i\gamma^0 mt)$, so the original has expectation

$$\langle E \rangle = \int d^3x \, \psi^\dagger \left(i \frac{\partial}{\partial t} \right) \psi = m \int d^3x \, \psi^\dagger \gamma^0 \psi . \tag{A26}$$

Replacing $\psi(x)$ with $\gamma^0 \psi(\tilde{x})$ changes the expectation to

$$\langle E \rangle = \int d^3x \, \psi^\dagger \gamma^{0\dagger} \left(i \frac{\partial}{\partial t} \right) \gamma^0 \psi = m \int d^3x \, \psi^\dagger \gamma^0 \psi . \tag{A27}$$

The components of the bi-spinor are chosen so that the mass and energy are positive in (A26). And whatever that choice, positivity of mass and energy are preserved in the parity-inverted case P_+. The parity inverted solution corresponding to P_- varies as $\exp(i\gamma^0 mt)$. Its stationary energy is therefore

$$
\begin{aligned}
\langle E \rangle &= \int d^3x \left(\gamma^2 \gamma^5 \gamma^0 \psi^*(\tilde{x}) \right)^\dagger \left(i\frac{\partial}{\partial t} \right) \left(\gamma^2 \gamma^5 \gamma^0 \psi^*(\tilde{x}) \right) \\
&= -\int d^3x\, \psi^T(\tilde{x}) \gamma^0 \gamma^5 \gamma^2 \left(i\frac{\partial}{\partial t} \right) \left(\gamma^2 \gamma^5 \gamma^0 \psi^*(\tilde{x}) \right) \\
&= \int d^3x\, \psi^T(\tilde{x}) \left(i\frac{\partial}{\partial t} \right) \psi^*(\tilde{x}) \\
&= -\left(\int d^3x\, \psi^\dagger(\tilde{x}) \left(i\frac{\partial}{\partial t} \right) \psi(\tilde{x}) \right)^*
\end{aligned}
\tag{A28}
$$

This energy is negative if $\psi(x)$ has positive energy. In any case, its energy is negated with respect to that of $\psi(x)$.

A.5 Time reversal
From (A8) and (A9),

$$
T_+ \left[\gamma^\mu \left(-i\partial_\mu + eA_\mu(x) \right) + m \right] = \gamma^{\mu\dagger} \left(i\partial_\mu + eA_\mu(-\tilde{x}) \right) + m .
\tag{A29}
$$

Taking the complex conjugate:

$$
\begin{aligned}
&\left[\gamma^{\mu\dagger} \left(i\partial_\mu + eA_\mu(-\tilde{x}) \right) + m \right]^* \left[T_+ \left[\psi(x) \right] \right]^* = 0 \\
&\Rightarrow \left(\gamma^{\mu T} \left(-i\partial_\mu + eA_\mu(-\tilde{x}) \right) + m \right) \left[T_+ \left[\psi(x) \right] \right]^* = 0
\end{aligned}
\tag{A30}
$$

Since $\gamma^{\mu T} = C\gamma^5 \gamma^\mu \gamma^5 C^{-1}$ then

$$
\gamma^5 C^{-1} \left[T_+ \left[\psi(x) \right] \right]^* = \eta \psi(-\tilde{x}) \Rightarrow T_+ \left[\psi(x) \right] = \eta^* C^{-1*} \gamma^{5*} \psi^*(-\tilde{x}) .
\tag{A31}
$$

It follows that T_+ is not in general a symmetry of the coupled Dirac equation, unless perhaps $A_\mu(-\tilde{x}) = A_\mu(x)$, though it is a *local symmetry* of the free Dirac particle equation according to (A31). In the Dirac representation this is

$$
T_+ \left[\psi(x) \right] = \gamma^2 \gamma^5 \gamma^0 \psi^*(-\tilde{x})
\tag{A32}
$$

(with η chosen so $T_+^2 = 1$). Applying (A21), the conjugate operator is

$$
T_- \left[\psi(x) \right] = \gamma^0 \psi(-\tilde{x}) .
\tag{A33}
$$

Following the steps in the section on parity inversion, the energy of a $\mathrm{T}_+\left[\psi(x)\right]$ solution (staying with the Dirac representation) is

$$
\begin{aligned}
\langle E \rangle &= \int d^3x \left(\gamma^5 \gamma^0 \gamma^2 \psi^*(-\tilde{x}) \right)^\dagger \left(i \frac{\partial}{\partial t} \right) \gamma^5 \gamma^0 \gamma^2 \psi^*(-\tilde{x}) \\
&= i \int d^3x\, \psi^T(-\tilde{x})\, \gamma^{2\dagger} \gamma^{0\dagger} \gamma^{5\dagger} \gamma^5 \gamma^0 \gamma^2 \frac{\partial \psi^*(-\tilde{x})}{\partial t} \\
&= i \int d^3x\, \psi^T(-\tilde{x}) \frac{\partial \psi^*(-\tilde{x})}{\partial t} \\
&= -i \int d^3x\, \psi^\dagger(-\tilde{x}) \frac{\partial \psi(-\tilde{x})}{\partial t}
\end{aligned}
\tag{A34}
$$

If again the original solution varies as $e^{-i\gamma^0 mt}$ then that will have energy given by (A26) whilst $\psi(-\tilde{x}) \sim e^{i\gamma^0 mt}$ which in (A34) gives

$$
\langle E \rangle = m \int d^3x\, \psi^\dagger \gamma^0 \psi
\tag{A35}
$$

so that the energy of the T_+ operation remains positive. For the $\mathrm{T}_-\left[\psi(x)\right] = \gamma^0 \psi(-\tilde{x})$ solution the energy is negative due to the change in sign of the exponent.

A.6 Charge conjugation

Charge makes its appearance in the Dirac equation only through the term $e\gamma^\mu A_\mu(x)$. If charge conjugation is applied to both the vector potential according to (A8) and the charge e of the particle according to Table A1 represented by the wavefunction, then this term and therefore the Dirac equation as a whole is unaffected by inversion; charge conjugation is a systematic symmetry of QED.

More interesting is if, for a wavefunction solving Dirac's equation for a given potential (sourced by fixed charges), there exists another simply-related wavefunction associated with a oppositely-charged particle. In this case

$$
\mathrm{c}\left[A_\mu(x)\right] = A_\mu(x), \quad \mathrm{c}\left[eA_\mu(x)\right] = -eA_\mu(x)
\tag{A36}
$$

and then

$$
\mathrm{c}_+\left[\gamma^\mu\left(-i\partial_\mu + eA_\mu(x)\right) + m\right] = -\gamma^\mu\left(i\partial_\mu + eA_\mu(x)\right) + m.
\tag{A37}
$$

Charge conjugation is a symmetry if there exists a $\psi_c(x) = \mathrm{c}_+\left[\psi(x)\right]$ that solves

$$
\left(-\gamma^\mu\left(i\partial_\mu + eA_\mu(x)\right) + m\right)\psi_c(x) = 0.
\tag{A38}
$$

Taking the complex conjugate:

$$
\left(\gamma^{\mu*}\left(i\partial_\mu - eA_\mu(x)\right) + m\right)\psi_c^*(x) = 0.
\tag{A39}
$$

Using (A17) this is

$$\left(\gamma^0 C\gamma^\mu C^{-1}\gamma^0\left(-i\partial_\mu + eA_\mu(x)\right) + m\right)\psi_c^*(x) = 0 \,.$$ (A40)

This is (A1) provided

$$\psi(x) = C^{-1}\gamma^0\psi_c^*(x) \Rightarrow \psi_c(x) = {\rm C}_+\big[\psi(x)\big] = \left(\gamma^0 C\psi(x)\right)^* \,.$$ (A41)

In the Dirac representation:

$${\rm C}_+\big[\psi(x)\big] = \left(i\gamma^0\gamma^2\gamma^0\psi(x)\right)^* = \left(-i\gamma^2\psi(x)\right)^* = -i\gamma^2\psi^*(x) \,.$$ (A42)

From (A19)

$${\rm C}_-\big[\psi(x)\big] = \gamma^{0*}C^*\gamma^{5*}\gamma^0 C\psi(x)$$ (A43)

which, in the Dirac representation, is

$${\rm C}_-\big[\psi(x)\big] = -\gamma^0 C\gamma^5\gamma^0 C\psi(x) = -\gamma^0\gamma^2\gamma^0\gamma^5\gamma^0\gamma^2\gamma^0\psi(x) = -\gamma^5\psi(x) \,.$$ (A44)

The energy of ${\rm C}_+\big[\psi(x)\big]$ is

$$\begin{aligned}
\langle E\rangle &= \int d^3x \left(-i\gamma^2\psi^*(x)\right)^\dagger\left(i\frac{\partial}{\partial t}\right)\left(-i\gamma^2\psi^*(x)\right)\\
&= \int d^3x\,\psi^T(x)\gamma^{2\dagger}\gamma^2\left(i\frac{\partial}{\partial t}\right)\psi^*(x)\\
&= -\left(\int d^3x\,\psi^\dagger(x)\left(i\frac{\partial}{\partial t}\right)\psi(x)\right)^*
\end{aligned}$$ (A45)

which is the negative of the original state. The energy of ${\rm C}_-\big[\psi(x)\big]$ is clearly positive.

A.7 Mass negation
With the action as defined in Table A1, applying (A10) and using (A8) and (A9),

$${\rm M}_+\big[\gamma^\mu\left(-i\partial_\mu + eA_\mu(x)\right) + m\big] = \gamma^\mu\left(-i\partial_\mu + eA_\mu(x)\right) - m = -\left(\gamma^5\gamma^\mu\gamma^5\left(-i\partial_\mu + eA_\mu(x)\right) + m\right)$$ (A46)

one immediately has

$${\rm M}_+\big[\psi(x)\big] = -\gamma^5\psi(x) = {\rm C}_-\big[\psi(x)\big] \,.$$ (A47)

It follows that

$${\rm M}_-\big[\psi(x)\big] = {\rm C}_+\big[\psi(x)\big] = \left(\gamma^0 C\psi(x)\right)^* \,.$$ (A48)

Notice that ${\rm M}_\pm = {\rm C}_\mp$ is to be expected from a cursory consideration of their effect on the Dirac equation. The energies however are not the same as for the equivalent charge

conjugations. If $\psi(x)$ is a solution with rest mass m, then $\gamma^5\psi(x)$ is a solution of the same equation but with rest mass $-m$. I.E.

$$\psi(x;-m) = -\gamma^5\psi(x;m). \tag{A49}$$

The effects of all the inversions are summarized in Table A2.

	i	$\langle E\rangle$	e	m	x	$A^\mu(x)$	constraints on EM coupling for point symmetry to exist	constraints on EM coupling for local symmetry to exist	symmetric partner of $\psi(x)$	Dirac representation
P_+	+	+							$\gamma^0\psi(\tilde{x})$	$\gamma^0\psi(\tilde{x})$
P_-	−	−	+	+	$\tilde{x}$	$h^\mu_\nu A^\nu(\tilde{x})$	$A^\mu(\tilde{x})=A^\mu(x)$	$A^\mu(x)\to A^\mu(t)$	$\left(\gamma^0 C\gamma^5\gamma^0\psi(\tilde{x})\right)^*$	$\gamma^2\gamma^5\gamma^0\psi^*(\tilde{x})$
T_+	+	+							$C^{-1}\gamma^{5*}\psi^*(-\tilde{x})$	$\gamma^2\gamma^5\gamma^0\psi^*(-\tilde{x})$
T_-	−	−	+	+	$-\tilde{x}$	$h^\mu_\nu A^\nu(-\tilde{x})$	$A^\mu(-\tilde{x})=A^\mu(x)$	$A^\mu(x)\to A^\mu(\mathbf{x})$	$\gamma^0\psi(-\tilde{x})$	$\gamma^0\psi(-\tilde{x})$
C_+	+	−					A^μ held fixed (not negated)	A^μ held fixed (not negated)	$\left(\gamma^0 C\psi(x)\right)^*$	$-i\gamma^2\psi^*(x)$
C_-	−	+	−	+	x	$-A^\mu(x)$			$\gamma^{0*}C^*\gamma^{5*}\gamma^0 C\psi(x)$	$-\gamma^5\psi(x)$
M_+	+	−							$-\gamma^5\psi(x)$	$-\gamma^5\psi(x)$
M_-	−	+	+	−	x	$A^\mu(x)$	none	none	$\left(\gamma^0 C\psi(x)\right)^*$	$-i\gamma^2\psi^*(x)$

Table A2. Inversions in Minkowski Spacetime.

Notes: The constraints on the coupling in the case of parity and time inversion are always satisfied if $A^\mu=0^\mu$ - i.e. there is no EM coupling. The symmetric partner under charge conjugation is the wavefunction of an oppositely charge particle (to that associated with the original solution $\psi(x)$) in the same field – i.e. without changing the sign of the charge of the sources of that field. Charge conjugation is a systematic symmetry of QED, in which case $\psi(x)$ is unaffected.

12. Acknowledgements

13. References

Abramowitz, M. & Stegun, I.A. (1965). *Handbook of Mathematical Functions*, Dover Publications, ISBN 9780486612720.

Barut, A. O. & Duru, I. H. (1987). Exact solutions of the Dirac equation in spatially flat Robertson-Walker space-times. *Physcial Review D*, 36, pp. 3705-3711.

Barut, A. O. & Singh, L. P. (1995). Dirac Equation in Spatially Nonflat Robetson-Walker Space-Times: II. *International Journal of Theoretical Physics*, 4, pp. 479-490.

Birrell, N. D. & Davies, P. C. W. (1982). *Quantum fields in curved space*, Cambridge University Press, ISBN 0 521 27858 9, Cambridge.

Calabi, E. & Marcus, L. (1962). Relativistic space forms. *Annals of Mathematics*, 75, pp. 63-76.

Coaescu, I. I. & Crusean, C. (2008). New Dirac quantum modes in moving frames of the de Sitter spacetime. *International Journal of Modern Physics*, A23, pp. 3707-3720.

Finster, F. & Reintjes, M. (2009). The Dirac equation and the normalization of its solutions in a closed Friedmann–Robertson–Walker universe. *Class.Quantum Gravity*, 26, pp. 105021-1-105021-20.

Friedrich, H. (2002). Conformal Einstein Evolution, In: The Conformal Structure of Spacetime Frauendiener, J. & Friedrich, H. Springer-Verlag, ISBN 3-540-44280-4, Berlin.

Hawking, S. W. & Ellis, G.F.R. (1973). Exact Solutions, In: The Large Scale Structure of Space-Time Cambridge University Press, ISBN 052109906-04, Cambridge.

Huang, X.-B. (2005). Exact solutions of the Dirac equation in Robertson-Walker space-time. http://arxiv.org/abs/gr-qc/0501077v1.

Ibison, M. (2007). On the conformal forms of the Robertson-Walker metric. *Journal of Mathematical Physics*, 48, pp. 122501-1-122501-23.

Ibison, M. (2010). Electrodynamics with a Future Conformal Horizon. pp. 28-42.

Itzykson, C. & Zuber, J.-B. (1985). *Quantum field theory*, McGraw-Hill, ISBN 0-07-066353-X, New York.

Kovalyov, M. & Légaré, M. (1989). The Dirac Equation in Robertson-Walker spaces: A class of solutions. *Journal of Mathematical Physics*, 31, pp. 191-198.

Lasenby, A. (2002). Conformal Geometry and the Universe. http://www.mrao.cam.ac.uk/~clifford/publications/abstracts/anl_ima2002. html.

Lasenby, A. & Doran, C. (2005). Closed Universes, de Sitter Space and Inflation. *Physical Review D*, 71, pp. 063502.

Nakamura, K. et al (2010). (Particle Data Group). *Journal of Physics G*, 37, pp. 075021.

Parashar, D. (1991). Dirac Equation in Robertson-Walker Metric. *International Journal of Theoretical Physics*, 30, pp. 1691-1695.

Penrose, R. (1963). Asymptotic properties of fields and space-time. *Physical Review Letters*, 10, pp. 66-68.

Peskin, M. E. & Schroeder, D. V. (1995). *An Introduction to Quantum Field Theory*, Addison-Wesley, ISBN 0 201 50397 2, Reading, Massachusetts.

Swaters, R. A., Sanders, R. H., and McGaugh, S. S. (2010). Testing Modified Newtonian Dynamics with Rotation Curves of Dwarf and Low Surface Brightness Galaxies. *The Astrophysical Journal*, 718, pp. 380-391.

Tomaschitz, R. (1994). Cosmological CP violation. *Journal of Mathematical Physics*, 35, pp. 1573-1596.

Villalba, V. M. & Percoco, U. (1989). *Journal of Mathematical Physics*, 31, pp. 715-720.

Weinberg, S. (1972). *Gravitation and Cosmology*, John Wiley & Sons, Inc., ISBN 0-471-92567-5, New York.

Weinberg, S. (2005). *The Quantum Theory of Fields, Volume 1: Foundations*, Cambridge University Press, ISBN 9780521670531, Cambridge, UK.

Part 4

Observational Tools

Statistical Study of the Galaxy Distribution

Antoine Labatie, Jean-Luc Starck and Marc Lachièze-Rey

Laboratoire AIM (UMR 7158), CEA/DSM-CNRS-Université Paris Diderot, IRFU, SEDI-SAP, Service d'Astrophysique, Centre de Saclay, F-91191 Gif-Sur-Yvette cedex France

1. Introduction

Large-scale structures in the Universe provide crucial information about formation of structures and can be used to test cosmological models. The good agreement between large-scale observations and the now-standard Lambda-Cold Dark Matter (ΛCDM) model gives hope for this model to be a lasting foundation. Going a step further, these observations have been used for precision cosmology to constrain cosmological parameters inside the model.

Large-scale structures are mainly studied through galaxy surveys, while on the other hand, cosmological models give predictions in terms of the global matter density field. Thus we need to understand the relationship between the matter density field and galaxy field, which is still a subject of research. Yet galaxy surveys have the advantage to map very large volumes. Recent galaxy surveys such as the Sloan Digital Sky Survey (SDSS, York et al. (2000)) and the 2 degree Field (2dF, Colless et al. (2001)) map unprecedented 3D volumes in the Universe. They have notably confirmed the view of large-scale structures as a collection of giant bubble-like voids separated by sheets and filaments of galaxies. This pattern has become known as the "Cosmic Web" (Bond et al. (1996), Springel et al. (2005)). The SDSS survey also enabled the first convincing detection of Baryon Acoustic Oscillations (BAOs) in large-scale structures (Eisenstein et al. (2005)), which opened a new field in precision cosmology.

Future galaxy surveys will map always bigger volumes, with more galaxies and continue to improve our knowledge of large-scale structures and cosmological models. Though data sets are of crucial importance, statistical methods extracting the information also have an important role. They should be optimized in order to exploit the full potential of these surveys. The purpose of this chapter is to review the different methods that can be used.

A first class of methods that we present is based on Fourier analysis, using the correlation function and the power spectrum. It originates from the simplicity of the Gaussian field, fully described by its second order moments. The Gaussian model gives a simple approximation of the matter density field, that works well on large scales. Another reason for the study of correlation function and power spectrum is that they are well understood and can be predicted in ΛCDM models.

Recently, Fourier analysis has been used to study BAOs. These structures are remnants of acoustic waves, traveling in the hot plasma before recombination, and that have stayed frozen in large-scale structures. The detection of BAOs in the SDSS has been a strong support for the ΛCDM model. Besides, they provide a statistical standard ruler since their absolute

size is known with small uncertainty. Thus their measurement in galaxy catalogues gives information on distances, and enables to constrain cosmological parameters.

A second class of methods use geometrical and topological tools to describe large-scale-structures. Among this class of methods, we present the Minkowski functionals and the genus statistic, providing a rigorous mathematical framework for the analysis. Among other advantages, Minkowski functionals are known analytically for gaussian random fields. Historically, the main application of those statistics was to determine the scale at which the distribution is gaussian, i.e. approximately the scale of linear evolution. Among other applications, we show how they can be used for model discrimination, or to provide a standard ruler in galaxy catalogues.

Finally we present the approach based on fractal analysis to characterize the galaxy distribution. It is motivated by the well-established scale-invariance of the galaxy clustering at small scales ($r < 10h^{-1}$ Mpc). Yet fractality on all scales would put into question current cosmological models, which assume large-scale homogeneity. It would also call into question the correlation function analysis, which assumes a well-defined mean density of the field. We discuss the question of the large-scale homogeneity. We show how the extension to multifractal analysis can represent more general distributions, in particular the distribution of galaxies, and enable to study the scale of homogeneity.

2. Fourier analysis of clustering

2.1 Power spectrum in the ΛCDM model

The Fourier analysis of clustering mainly originates from a simple approximation of the large-scale distribution of matter being a Gaussian field. The justification is that, in current cosmological models involving inflation, the matter distribution originates from tiny gaussian random fluctuations that gravitationally evolved. Let us note δ the fluctuation field, from the matter density field ρ and its mean value $\bar{\rho}$:

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \bar{\rho}}{\bar{\rho}} \tag{1}$$

In the linear regime of evolution, the fluctuation is supposed to be small ($\delta \ll 1$), and the equation of evolution is linear with δ:

$$\frac{\partial^2 \delta}{\partial t^2} + 2\frac{\dot{a}}{a}\frac{\partial \delta}{\partial t} = 4\pi G\bar{\rho}\delta \tag{2}$$

with a the expansion factor of the Universe, $\dot{a}$ its time derivative (i.e. $H = \frac{\dot{a}}{a}$), G Newton's gravitational constant.

In this regime, the Fourier modes $\tilde{\delta}(\mathbf{k})$ all evolve proportionally to an evolution factor. There are two solutions for this equation of evolution, with one growing mode and one decaying mode. The growing mode, which necessarily governs the evolution, has the following expression in a ΛCDM model:

$$G(z) = \frac{5\Omega_m E(z)}{2} \int_z^\infty \frac{(1+z')}{E(z')^3} dz' \tag{3}$$

with Ω_m the dimensionless matter density and $E(z) = \frac{H(z)}{H_0}$ the dimensionless Hubble constant at redshift z. In the linear regime, the gaussianity of the field is preserved, as well as

the gaussianity of the Fourier coefficients $\tilde{\delta}(\mathbf{k})$ of the fluctuation. Thus a complete description of the field can be given in this regime in terms of the power spectrum $P(\mathbf{k})$ of the fluctuation field

$$\mathbb{E}[\tilde{\delta}(\mathbf{k})\tilde{\delta}^{\star}(\mathbf{k}')] = (2\pi)^3 \delta_D(\mathbf{k} - \mathbf{k}')P(\mathbf{k}) \tag{4}$$

with δ_D the 3-dimensional Dirac function. The power spectrum quantifies the expected strength of the oscillations at wavelength $\mathbf{k}$. With the usual hypothesis of isotropy, the power spectrum only depends on the norm k of the vector $\mathbf{k}$. However in galaxy catalogues, the galaxy are studied in redshift space which slightly breaks the isotropy property.

The linear approximation works very well at large scale, where the strength of the fluctuation is small. On smaller scale, the matter field is in a much more advanced evolution phase and has already overtaken non-linear evolution, which makes it much harder to understand. To take into account non-linear evolution, one has to use N-body simulations and empirically infer formulas.

Overall, the dark matter power spectrum can be written as

$$P_{DM}(k,z) = G^2(z)T^2(k,z)P_l(k) \tag{5}$$

where $T(k,z)$ is the transfer function at redshift z including non-linear evolution and $P_l(k)$ is the primordial dark matter power spectrum.

2.2 Two-point correlation function in the ΛCDM model

Just as the power spectrum, the two-point correlation function is a second order statistic that describes the clustering of a field or a point process. It is simply the inverse Fourier transform of the power spectrum, i.e. when P and ξ are isotropic:

$$\xi(r) = \frac{1}{2\pi^2} \int k^2 P(k) \frac{\sin(kr)}{kr} dk \tag{6}$$

The correlation function $\xi(r)$ at distance r also has a statistical interpretation. It measures the excess of probability to find a pair of points in two volumes dV_1 and dV_2 at distance r compared to a random distribution.

$$dP_{12} = \bar{\rho}^2[1 + \xi(r)]dV_1 dV_2 \tag{7}$$

The correlation function is directly linked to the fluctuation field $\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \bar{\rho}}{\bar{\rho}}$:

$$\xi(r) = \langle \delta(\mathbf{x})\delta(\mathbf{x} + \mathbf{r}) \rangle \tag{8}$$

with $\|\mathbf{r}\| = r$. As for the power spectrum analysis, the standard ΛCDM model predicts a certain shape for $\xi(r)$ with a dependence on the cosmic parameters. In this model the correlation is rapidly decreasing at small scale as a power law as found in the distribution of galaxies (Norberg & Baugh (2001), Zehavi et al. (2005)):

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma} \tag{9}$$

with $r_0 \approx 5h^{-1}$Mpc and $\gamma \approx 1.7$.

At large scale the correlation presents a very interesting feature caused by BAOs. It is a relic of sound waves traveling in the early Universe when baryon and photons were coupled in a

relativistic plasma before recombination which caused the wave propagation to end Eisenstein & Hu (1998). It has imprinted the large scale distribution and can be seen as a small peak in the correlation function at a scale r_s, corresponding to the comoving distance of the sound horizon.

2.3 Modeling clustering bias in the galaxy distribution

As described in the previous section, theoretical predictions about clustering are given for the global dark matter density field. On the other hand, current large-scale surveys map the galaxy field. Thus, if we want to relate cosmological predictions to these observations, we have to understand the link between dark matter and galaxies.

The distribution of galaxies n_g is usually modeled by a Cox process, i.e. a Poisson process with an intensity given by a continuous field $\rho_g(\mathbf{x})$, which itself is a statistical process. Knowing $\rho_g(\mathbf{x})$, the number of galaxies in a volume dV around $\mathbf{x}$ is a Poisson variable with intensity $\rho_g(\mathbf{x})dV$. The field ρ_g is linked to the underlying dark matter density field ρ_m (at least stochastically) since galaxies tend to form in over-densities, but is not supposed to be identical. As a consequence, there are differences in the clustering of the two fields, and thus in their correlation functions $\xi_g(r)$ and $\xi_m(r)$. The relationship between ρ_g and ρ_m depends on the way that galaxies are grafted onto the matter density field. This is a critical question that remains an active research subject.

Historically, it has been observed that clusters of galaxies cluster more strongly than galaxy themselves. In more recent surveys, more luminous galaxies have been found to cluster more strongly. This shows a non trivial relation between the dark matter and the visible matter, which depends also on the type of objects. This effect is referred as the bias of visible matter with respect to the dark matter:

$$\xi_g(r) = b(r)^2 \, \xi_m(r) \tag{10}$$

with $b(r)$ the bias of the galaxy field with respect to dark matter. The bias b modulates the galaxy correlation with respect to the dark matter correlation and should depend on the scale. This concept of bias appeared in the 80's to reconcile the CDM model with observational constraints. A first attempt to understand this bias was made by Kaiser in 1984 with the high peak model, which considers the formation of objects at local maximums of the smoothed density field.

Recently, the halo model (Cooray & Sheth (2002)) has been a fruitful step in order to connect the galaxy and dark matter distributions. A given halo model predicts the expected number of galaxies above given luminosity thresholds inside halos of different mass. Halo models can be parametrized to match very well current observations and naturally explains the biasing phenomenon. The halo model has become particularly useful in cosmological N-body simulations, where it is used to populate dark matter halos with galaxies, providing realistic simulations for current and future galaxy surveys.

2.4 Correlation function estimation from galaxy surveys

Both the correlation function and power spectrum have been extensively measured on galaxy surveys. The power spectrum has the advantage that the different bands are weakly correlated, which is not the case for the correlation function. On the other hand the correlation function is more intuitively linked to 3D space. As an illustration, BAOs from the early plasma stayed frozen at the sound horizon scale r_s, which produces an excess of clustering at this

particular distance. Thus BAOs manifest as a localized bump at the sound horizon scale in the correlation function, whereas they manifest as an infinite number of wiggles in the power spectrum.

In this section we decide to focus on the correlation function, explaining details about its estimation and how it can be used for BAO study.

2.4.1 Correlation function estimation

When estimating the correlation function, there are two types of statistical uncertainties. The first one comes from cosmic fluctuations due to limited sample volume, and the other one from the finite number of galaxies which do not trace exactly the underlying field (i.e. shot noise). In general, estimators of the correlation are subject both to bias and variance. In practice there is no way to evaluate the bias of the estimator if it exists, and it must be considered itself as a source of uncertainty, in addition to the variance.

There are various estimators of the correlation function, most using random catalogues with identical geometry to measure this excess of probability. Let N_D and N_R be the number of galaxies respectively in the data and random catalogues. We define $DD(r)$, $RR(r)$ and $DR(r)$ as the number of pairs at a distance in $[r \pm dr/2]$ of respectively data-data, random-random and data-random galaxies. We also define N_{DD}, N_{RR} and N_{DR} as the total number of corresponding pairs in the (real or random) catalog. With the convention of counting pairs only once we have:

$$N_{DD} = \frac{N_D(N_D - 1)}{2} \tag{11}$$

$$N_{RR} = \frac{N_R(N_R - 1)}{2} \tag{12}$$

$$N_{DR} = N_R N_D \tag{13}$$

$$\tag{14}$$

Let us give the expression of 4 of the most commonly used estimators used, from Peebles & Hauser (1974), Davis & Peebles (1983), Hamilton (1993) and Landy & Szalay (1993)

$$\hat{\xi}_{PH}(r) = \frac{N_{RR}}{N_{DD}} \frac{DD(r)}{RR(r)} - 1$$

$$\hat{\xi}_{DP}(r) = \frac{N_{DR}}{N_{DD}} \frac{DD(r)}{DR(r)} - 1$$

$$\hat{\xi}_{HAM}(r) = \frac{N_{DR}^2}{N_{DD}N_{RR}} \frac{DD(r)RR(r)}{[DR(r)]^2} - 1$$

$$\hat{\xi}_{LS}(r) = 1 + \frac{N_{RR}}{N_{DD}} \frac{DD(r)}{RR(r)} - 2\frac{N_{RR}}{N_{DR}} \frac{DR(r)}{RR(r)}$$

In general, the analysis is complicated by selection effects in the catalogues, with a density decreasing farther away. One has to consider a mean density varying with the position in the catalogue $\bar{\rho}(\mathbf{x})$, i.e. fluctuations $\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \bar{\rho}(\mathbf{x})}{\bar{\rho}(\mathbf{x})}$. This is taken into account in the random catalogues which have the same selection effect as the catalogue. Finally an additional weight

can be given to the galaxies at different position of the survey (both in the data and random catalogues) in order to minimize the estimator's variances (Kaiser (1986)).

Estimating ζ would be easier knowing the exact number of points in the volume expected from the distribution. In practice we can only estimate it with the empirical quantities N_D and N_{DD}. Let us show that Hamilton and Landy-Szalay only depend on the second order on this uncertainty in the mean density, and thus perform better. We repeat the calculations given in Hamilton (1993) in a simple case where the sample is volume-limited (i.e. with a constant expected density $\bar{\rho}$ in the sample), so that the optimal strategy is to weight all galaxies equally. The empirical density in the catalogue ρ is a sum of Dirac functions on the galaxies of the catalogue. For a volume-limited survey with constant expected density $\bar{\rho}$, the relative fluctuation in the sample is given by

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \bar{\rho}}{\bar{\rho}} \tag{15}$$

We write W the indicator function of the sample volume and $\langle . \rangle$ the integration on the volume. For example $\langle W(\mathbf{x}) \rho(\mathbf{x}) \rangle$ is the integration of the empirical density and thus equals the number of points in the sample. We introduce the following quantities that have statistical expectations of 0 for $\bar{\delta}$ and Ψ, and ζ for $\hat{\zeta}$:

$$\bar{\delta} = \frac{\langle W(\mathbf{x}) \, \delta(\mathbf{x}) \rangle}{\langle W(\mathbf{x}) \rangle} \tag{16}$$

$$\Psi(r) = \frac{\langle \delta(\mathbf{x}) \, W(\mathbf{x}) W(\mathbf{y}) \rangle_r}{\langle W(\mathbf{x}) W(\mathbf{y}) \rangle_r} \tag{17}$$

$$\hat{\zeta}(r) = \frac{\langle \delta(\mathbf{x}) \delta(\mathbf{y}) \, W(\mathbf{x}) W(\mathbf{y}) \rangle_r}{\langle W(\mathbf{x}) W(\mathbf{y}) \rangle_r} \tag{18}$$

where $\langle . \rangle_r$ means a double integration in the volume, restricted to $\mathbf{x}$ and $\mathbf{y}$ separated by a distance in $[r \pm dr/2]$. $\hat{\zeta}$ is an unbiased estimator of the real ζ but we cannot calculate it since we do not know $\bar{\rho}$ and δ.

With short calculations it is possible to express the different estimators with the quantities $\hat{\zeta}$, $\bar{\delta}$ and Ψ (Hamilton (1993)):

$$\hat{\zeta}_{PH}(r) = \frac{\hat{\zeta}(r) + 2\Psi(r) - 2\bar{\delta} - \bar{\delta}^2}{[1 + \bar{\delta}]^2} \tag{19}$$

$$\hat{\zeta}_{DP}(r) = \frac{\hat{\zeta}(r) + \Psi(r) - \bar{\delta} - \Psi(r) \, \bar{\delta}}{[1 + \bar{\delta}] \, [1 + \Psi(r)]} \tag{20}$$

$$\hat{\zeta}_H(r) = \frac{\hat{\zeta}(r) - \Psi(r)^2}{[1 + \Psi(r)]^2} \tag{21}$$

$$\hat{\zeta}_{LS}(r) = \frac{\hat{\zeta}(r) - 2\bar{\delta}\Psi(r) + \bar{\delta}^2}{[1 + \bar{\delta}]^2} \tag{22}$$

These formulas explain the superiority of Hamilton and Landy-Szalay estimators, with Ψ and $\bar{\delta}$ terms at the second order in the numerator. Terms in the denominator are not important since they generate a small relative error, whereas terms in the numerator can generate a high relative error when their values become non negligible compared to $\hat{\zeta}$. For Hamilton and

Landy-Szalay estimators, the error is dominated by the one of $\hat{\bar{\xi}}$ and not really affected by Ψ and $\bar{\bar{\delta}}$, which are linked to the uncertainty in $\bar{n}$.

The superiority of Hamilton and Landy-Szalay has been observed empirically in different studies (Labatie et al. (2010), Kerscher et al. (2000), Pons-Bordería et al. (1999)). In Labatie et al. (2010) it is also shown that these estimators are nearly unbiased for current galaxy surveys. In particular the integral constraint (caused by the uncertainty in the mean density $\bar{\rho}$) has nearly no effect on the estimation. We show in figure 1 results with different estimators when estimating the correlation function on simulations of the SDSS Luminous Red Galaxies (LRG) sample. Estimators are nearly unbiased for this survey size and Hamilton and Landy Szalay estimators show much less variance than the other estimators.
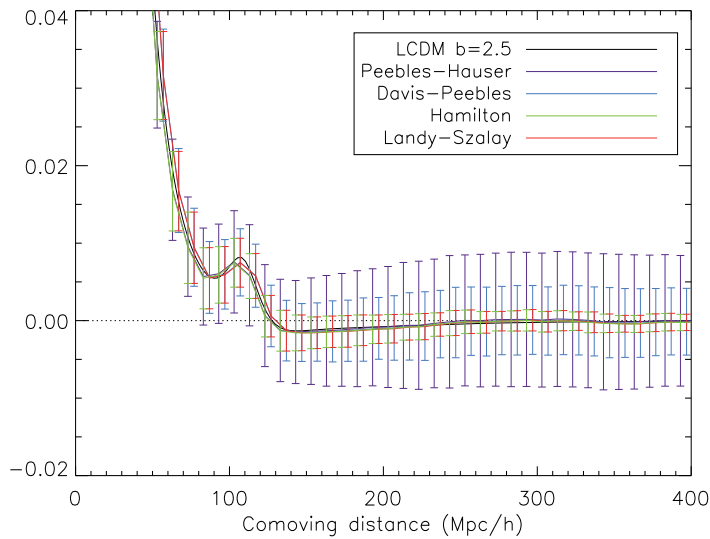

Fig. 1. Estimation of the correlation function on the SDSS LRG catalogue. The different colored curves show the result for each estimator with error bars showing the standard deviations of the estimation. They were estimated using 2000 simulations. The real correlation function is given by the black curve. The small bump in the correlation at scale $r \approx 105h^{-1}$Mpc corresponds to the BAO feature. The bias is negligible for all estimators but there are great differences in their quality. Bias and uncertainty are important to know the error when fitting the estimated correlation function to a model correlation function (as in BAO studies).

As a result, recent studies usually focus on the Landy-Szalay estimator to measure the galaxy correlation function.

2.4.2 Correlation function and BAO study

Recently a lot of studies using the power spectrum or the correlation function have been motivated by BAO study. The reason is that BAOs provide a statistical standard ruler, i.e. a powerful tool to study the geometry of the Universe. Since their absolute size is known with small uncertainty from measurements in the Cosmic Microwave Background (CMB), their detection in galaxy surveys gives access to real distances in the catalogue.

The primary informations obtained on galaxies are the angular positions of the objects on the sky, and their redshift. In order to obtain 3D positions, one has to assume a fiducial cosmological model and convert redshifts into distances. We show in figure 2 how the size of an object is related to its angular extent θ and its redshift extent dz, using the real cosmological quantities $d_A(z)$ (angular diameter distance) and $H(z)$. The interest of the standard ruler is that its size is already known, and thus from the angular and redshift extents, it naturally provides the quantities $d_A(z)$ and $H(z)$ at the redshift of the survey.
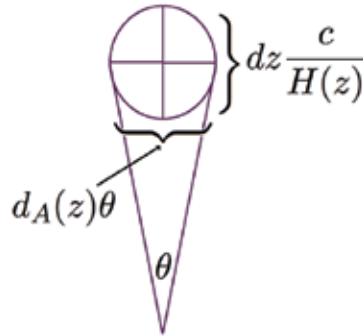

Fig. 2. Size of a spherical object as a function of its angular extent θ and redshift extent dz, using the real angular diameter distance $d_A(z)$ and Hubble constant $H(z)$ at redshift z of the survey. Figure adapted from Bassett & Hlozek (2010).

BAOs constrain cosmological parameters through $d_A(z)$ and $H(z)$ in a complementary way with CMB and type Ia supernovae, which explains the recent interest in this field. BAOs were first convincingly detected in the correlation function of the SDSS survey (Eisenstein et al. (2005)).

BAO study through correlation can be divided in two different parts. The first one consists in detecting the BAO feature in the correlation function, which can be done by rejecting the null hypothesis that there are no BAOs. When the detection is made significantly, the second step is to determine distances in the survey, which provides the best fit in the BAO peak between the measured and the model correlation functions.

2.4.2.1 BAO detection and significance

Let us describe more precisely the procedure for the detection. In most BAO studies (for example Eisenstein et al. (2005), Hütsi (2006), Percival et al. (2010), Blake et al. (2011)), the χ^2 statistic is used. Here the observations are the estimated correlation function values in different bins $(\hat{\xi}_i)_{1 \leq i \leq n} = (\hat{\xi}(r_i))_{1 \leq i \leq n}$. Given the covariance matrix $(C_{i,j})$ of this measurement vector, and the expected value given by the model $\xi = (\xi_i)$, the χ^2 statistic is defined as

$$\chi^2 = (\hat{\xi} - \xi)^T C^{-1} (\hat{\xi} - \xi)$$
$$= \sum_{i,j} (\hat{\xi}_i - \xi_i) C_{i,j}^{-1} (\hat{\xi}_j - \xi_j)$$

The covariance matrix is usually estimated empirically using mock catalogues, and then inverted. Assuming that the measurement vector is gaussian, the χ^2 statistic follows a

$\chi^2(n)$ law, i.e. a chi-square law with n degrees of freedom. The situation is a bit more complex here because the model correlation function $\xi(\theta_1, ..., \theta_k)$ depends on cosmological parameters $(\theta_1, ..., \theta_k)$, and so the $\chi^2(\theta_1, ..., \theta_k)$ statistic also depends on a choice of model. The method used consists in finding the best fit model $(\theta_1^0, ..., \theta_k^0)$, i.e. the model with $\xi(\theta_1, ..., \theta_k)$ minimizing the χ^2 statistic

$$(\theta_1^0, ..., \theta_k^0) = \underset{(\theta_1, ..., \theta_k)}{\operatorname{argmin}} \chi^2(\theta_1, ..., \theta_k) \tag{23}$$

With k parameters used for the fitting, it can be shown that the χ^2 at the best fit value follows a $\chi^2(n-k)$ law (under the assumption that one of the model is correct), i.e. a chi-square law with $n-k$ degrees of freedom. Even though this result is widely used, it is worth noting that it is subject to regularity assumptions The best fit χ^2 statistic is computed with the 2 different models (with and without BAOs) which leads to a difference $\Delta\chi^2$ in the best fit χ^2 for the two different models.

This difference $\Delta\chi^2$ can be used to test between 2 competing model that are nested, i.e. one model can be obtained by fixing/eliminating parameters in the other model. If we note 1 the model with fewer parameters, i.e. a higher value of χ^2, and 2 the other model, then it can be shown that the difference $\chi_{\text{diff}} = \chi_1^2 - \chi_2^2$ follows a $\chi^2(d_{diff})$ law (under the assumption that the restrained model is also correct). Here, $d_{diff} = d_1 - d_2$ is the difference in the number of parameters of the two models. In the case we are interested in, an artificial parameter is introduced to account for the presence or not of BAOs in the correlation function. The global model then becomes an union of the 2 precedent models and the restrained model contains only the model without BAOs. Then the χ_{diff} corresponds to the difference $\Delta\chi^2$, with a value that can be interpreted in terms of significance.

This methodology has been extensively used, for example in Eisenstein et al. (2005) for the first BAO detection at the $3.4\,\sigma$ level (the convention is that a number of σ's is converted from a p-value using the relation for a Gaussian distribution). In Cabré & Gaztañaga (2011) it is used to show with mock catalogues that a detection higher than 3σ is not likely in the SDSS.

2.4.2.2 Constraining cosmological parameters through BAO scale

A later step consists in using the BAO feature in the correlation function to constrain cosmological parameters. We remind that 3D catalogues are created assuming a fiducial cosmology to obtain the distance from the observer to the objects. If we change the fiducial cosmology to the real cosmology, then the objects move and dilate. Spherical objects have a size which approximately scales as the dilation scale $D_V(z) = (1+z)^2 d_A(z)cz/H(z)$ (Eisenstein et al. (2005)) . Therefore, the correlation function in the true cosmology is obtained by dilating it with the ratio of the true to the fiducial dilation scales. When doing the χ^2 test this dilation scale is introduced as a parameter, and an estimation of the true dilation scale is obtained for the best fit. Other cosmological parameters can also be introduced during the fitting procedure. The dilation scale itself provides constraints since it depends on cosmological parameters.

In Eisenstein et al. (2005), there are 3 fitted parameters: an overall amplitude A, $\Omega_m h^2$ and $D_V(0.35)$, where $z = 0.35$ is the mean redshift of the SDSS LRG sample used. The other parameters intervening in the correlation function are kept fixed because they are well constrained by other measurements, e.g. on the CMB. The posterior probability on those parameters is simply equal to the likelihood (assuming a flat prior). This is also directly linked

to the $\chi^2(A, \Omega_m h^2, D_V(0.35))$ statistic with the gaussian hypothesis. So the χ^2 statistic directly gives the confidence regions for these parameters as shown in figure 3, where the predictions of different models are overplotted.
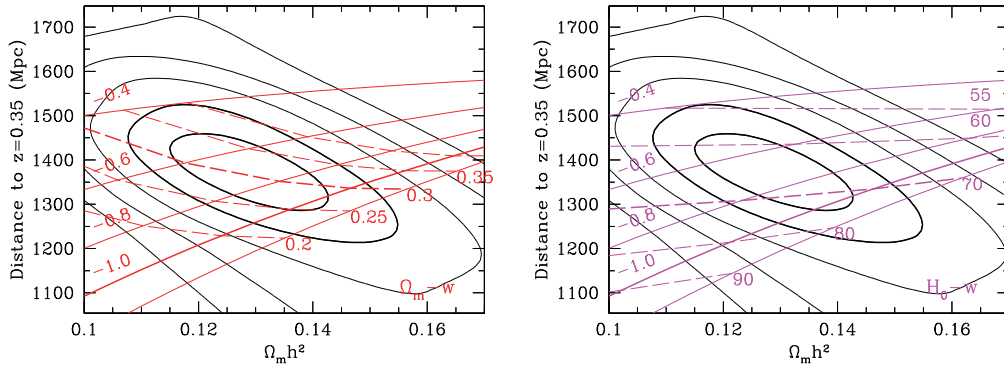
Baryon Acoustic Oscillations

Fig. 3. The likelihood contours of CDM models as a function of $\Omega_m h^2$ and $D_V(0.35)$. The likelihood has been taken to be proportional to $\exp(-\chi^2/2)$, and contours corresponding to 1σ to 5σ for a 2D Gaussian have been plotted. Overplotted are the model predictions for fixed Ω_m, w and H_0. Figure courtesy of Eisenstein et al. (2005).

3. Geometrical analysis of clustering

In the first section we presented the Fourier analysis of clustering based on the second order statistics which are the power spectrum and its real space equivalent, the 2-point correlation function. In general a random field is characterized by all its moments and N-point correlation functions. The 2-point correlation function necessarily presents limitations for model discrimination and may not be the best tool for some applications. One example of such application is the geometrical analysis of large-scale structures. Obviously the 2-point correlation function does not provide much geometrical information since it provides spherically averaged quantities.

In recent galaxy surveys, the Universe appears as a collection of giant bubble-like voids separated by sheets and filaments of galaxies, with the superclusters appearing as occasional relatively dense nodes. This network, usually called the "Cosmic Web", is clearly visible in the 2dF Galaxy Redshift Survey and the SDSS survey.

The study of the geometry and topology of large-scale structures can carry useful information and complement the study of second order clustering. We review different geometrical methods used for this task and results obtained.

3.1 Minkowski functionals

Minkowski functionals are a very useful tool to describe hypersurfaces of dimension $d-1$ in a space of dimension d. Hadwiger theorem states that any function defined on subsets A of the space that satisfies the properties of translation and rotation invariance, additivity and continuity can be decomposed as a sum of the Minkowski functionals. In a space of dimension

d there are $d + 1$ Minkowski functionals. In the case of the 3D distribution of galaxies, where $d = 3$, Minkowski functionals are

$$V_0(A) = \int_A d^3\mathbf{x}$$

$$V_1(A) = \frac{1}{6} \int_{\delta A} dS(\mathbf{x})$$

$$V_2(A) = \frac{1}{6\pi} \int_{\delta A} \left(\frac{1}{R_1(\mathbf{x})} + \frac{1}{R_2(\mathbf{x})} \right) dS(\mathbf{x})$$

$$V_3(A) = \frac{1}{4\pi} \int_{\delta A} \frac{1}{R_1(\mathbf{x})R_2(\mathbf{x})} dS(\mathbf{x})$$

where R_1 and R_2 are the principal curvatures of the surface δA.

These functions have intuitive interpretations in the case $d = 3$. V_0 corresponds to the volume of A, V_1 is proportional to the surface of the boundary δA, V_2 corresponds to the mean curvature on the surface δA and V_3 is half the Euler characteristic χ of the subset, linked to the genus G

$$\chi = 2(1 - G) \tag{24}$$

V_3 is a topological measure, which equals 1 for a ball and any analogue convex set. It can be interpreted as the surface drawn on the unit sphere by the vector normal to the hypersurface, when covering δA. The surface is counted positively when the product of the curvature radii is positive and negatively otherwise.

The genus also has a simple interpretation. It corresponds to the number of torus holes, minus the number of isolated regions, plus 1.

3.1.1 Characterization of random fields

Minkowski functionals can be calculated on scalar fields, for example on realizations of random fields or on the distribution of galaxies Gott et al. (1986). In this context Minkowski functionals are calculated on the excursion sets of these fields, i.e. on the isodensity surfaces. The excursion set of a function for a threshold u is given by the set $F_u = \{\mathbf{x}, Z(\mathbf{x}) \geq u\}$. It is more complicated for a point distribution like the galaxy distribution since excursion sets are not well defined. Thus, the point distribution is usually smoothed at different scales giving a continuous 3D field, where the Minkowski functionals can be calculated.

An example of isodensity surfaces is given in figure 4 for a Gaussian random field with a power spectrum $P(k) \sim k$ smoothed with kernel width $s = 3$ in a cube of size 128. The left column shows the lower density regions and the right column the higher density regions with respective fractions of the volume of 7%, 50% and 93%. The symmetry between high and low density is clearly seen. Another important aspect of Gaussian field is the sponginess at median threshold, quantified by a positive genus, which can be seen here and also in the matter distribution (Gott et al. (1986)).

For random realizations, Minkowski functionals have statistical variations as the field itself. However if the field is homogeneous above a certain scale (i.e. if it presents an ergodicity property), Minkowski functional densities (i.e. Minkowski functionals divided by the volume) converge to their expected value.
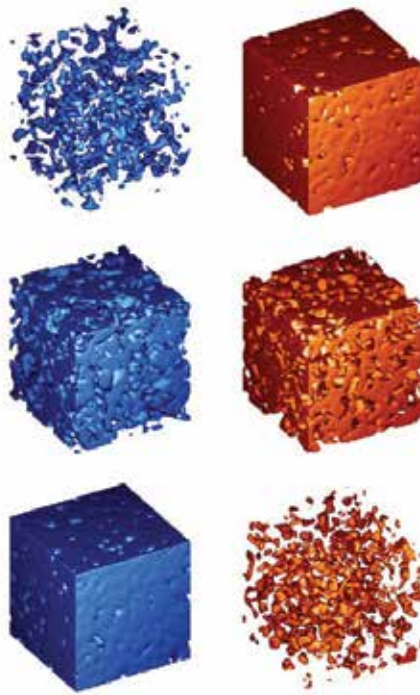
Fig. 4. Isodensity surface of a Gaussian random field smoothed with kernel width $s = 3$. The left column shows the lower density regions and the right column the higher density regions with respective volume fractions of 7%, 50% and 93%.The symmetry between high and low density is clearly seen, as well as the sponginess of the distribution at the median density threshold. Figure courtesy of Martínez et al. (2005).

A change of variable is often made to replace the threshold u by a standard variable, $v = \frac{u-\mu}{\sigma}$, where μ and σ are respectively the expected value and standard deviation of the field. Finally the Minkowski functionals are divided by the total volume studied, which gives functional densities which can be compared between different volumes. A great advantage of Minkowski functionals is that their values are known for Gaussian fields. The expected value of the first Minkowski functional $V_0(u)$ is given by the cumulative density function of a standard Gaussian distribution.

$$\mathbb{E}[V_0(u)] = V_{tot} \frac{1}{\sqrt{2\pi}\sigma} \int_u^\infty e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \tag{25}$$

with $v = \frac{u-\mu}{\sigma}$ this gives

$$\mathbb{E}[V_0(v)] = V_{tot} \frac{1}{\sqrt{2\pi}} \int_v^\infty e^{-\frac{t^2}{2}} dt \tag{26}$$

So the functional density is given by

$$\mathbb{E}[v_0(\nu)] = \frac{1}{\sqrt{2\pi}} \int_\nu^\infty e^{-\frac{t^2}{2}} \, dt \tag{27}$$

In Tomita (1990) the theoretical values of the other Minkowski functional densities are given in expected value

$$\mathbb{E}[v_1(\nu)] = \frac{2}{3} \frac{\lambda}{\sqrt{2\pi}} \exp\left(\frac{-\nu^2}{2}\right) \tag{28}$$

$$\mathbb{E}[v_2(\nu)] = \frac{2}{3} \frac{\lambda^2}{\sqrt{2\pi}} \nu \exp\left(\frac{-\nu^2}{2}\right) \tag{29}$$

$$\mathbb{E}[v_3(\nu)] = \frac{\lambda^3}{\sqrt{2\pi}} (\nu^2 - 1) \exp\left(\frac{-\nu^2}{2}\right) \tag{30}$$

where $\lambda = \sqrt{\frac{\sigma_1^2}{6\pi\sigma^2}}$, $\sigma = \mathbb{E}\left[\delta^2\right]^{1/2}$, $\sigma_1 = \mathbb{E}\left[|\nabla\delta|^2\right]^{1/2}$, and δ is the fluctuation field.

Note that for gaussian fields, Minkowski functionals are unique up to a multiplicative factor. This is very useful to test for gaussianity, which was the primary use of Minkowski functionals (in particular the genus). Since the matter distribution originates from a Gaussian field, it should also remain gaussian when smoothed at the linear scales of evolution. This overall picture can be tested by smoothing the field at the linear scales and comparing Minkowski functionals with the ones expected for a Gaussian field.

3.1.2 Results obtained on galaxy surveys with Minkowski functionals

3.1.2.1 Test of gaussianity

Let us first illustrate how to test the gaussianity of the field using the third Minkowski functional v_3, which is a topology measure. We perform the calculation for a catalogue described in Martínez et al. (2005). It is a volume limited sample extracted from the 2dF Galaxy Redshift Survey, which contains 8487 galaxies and has been cut to obtain a rectangular volume.

We show in figure 5 the $v_3(\nu)$ curve with a smoothing of the field by gaussian windows with different sizes $s = 1, 2, 4$ and $8\,h^{-1}$Mpc. The smoothing consists in applying a low-pass filter on the field, and only keeping the low Fourier modes. For $s = 4, 8\,h^{-1}$Mpc the remaining Fourier modes present a v_3 statistic very similar to a gaussian field. This suggests that non-linear evolution has not much changed the field topology at these scales. On the contrary, for $s = 1, 2\,h^{-1}$Mpc, there are significant deviations from a gaussian topology.

In Martínez et al. (2005) the galaxy field is first denoised using a multiscale wavelet decomposition. The idea is to look at the topology of the real continuous field, instead of the discrete galaxy distribution smoothed at different scales. Results show that the continuous field has clearly a non-gaussian topology, whereas the galaxy field becomes rapidly gaussian when increasing the smoothing size. Thus, even though the matter distribution is gaussian at linear scales, its real global topology is highly non-gaussian.

3.1.2.2 Model discrimination

A second application of Minkowski functionals, and in particular the genus, is the power of discrimination between models. In Sheth et al. (2003), Minkowski functionals are calculated
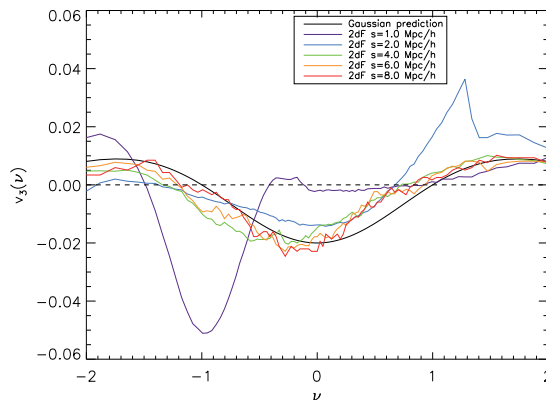
Fig. 5. $v_3(v)$ statistic on the 2dF survey. The different curves show $v_3(v)$ obtained when smoothing the galaxy field at different scales $s = 1, 2, 4, 6$ and $8\,h^{-1}$Mpc (color curves). We also show the $v_3(v)$ expected for a Gaussian field from equation (30) (black curve). Each curve has been globally rescaled to make the comparison possible. We see the curves approaching the gaussian result when increasing the smoothing radius. This supports the fact that the galaxy field becomes gaussian at large scale.

on cosmological simulations of the Virgo consortium, using SURFGEN. Calculations are made on 3 rival cosmological models - a ΛCDM model with $\Omega_m = 0.3$, a Standard-CDM (SCDM) model of a flat Universe with $\Omega_m = 1$, and a τCDM model also flat and with decaying massive τ neutrino as the dark matter constituent. The models show very distinct morphologies, which can be easily distinguished through their Minkowski functionals.

In Hikage et al. (2003) Minkowski functionals are calculated on the SDSS Early Data Release showing an agreement (up to statistical uncertainties) between estimated Minkowski functionals in galaxy catalogues and Minkowski functionals obtained from ΛCDM N-body simulations. Results also show a clear incompatibility with simulations in the SCDM model.

3.1.2.3 Constraining cosmological parameters

Minkowski functionals can also be used to infer real distances of the survey. As for BAOs, the topology of large-scale structures (quantified by the genus statistic) can provide a standard ruler. Up to statistical uncertainties, survey distances can be adjusted to obtain a measured genus that matches the one expected for a given cosmological model. This approach has been first proposed in Park & Kim (2010), and applied in Zunckel et al. (2011) to show how it could constrain the equation of state of the dark energy for the upcoming Baryon Oscillation Spectroscopic Survey (BOSS, www.sdss3.org). They show that BOSS could constrain a constant parameter w in the equation of state with a 5% uncertainty, using this approach.

The quality of a standard ruler depends on the precision in the knowledge of its real size, and also on the uncertainty of the measure. Here, the knowledge of the expected genus may not be as good as the knowledge of the baryon acoustic scale. Yet the genus has the advantage that it can be applied at different scales in the distribution, thus decreasing the measurement uncertainty. Therefore it could be very useful in order to cross-check results obtained using BAOs.

4. Fractal analysis

4.1 The concept of fractal

The idea of a fractal Universe is very old, and already in the eighteenth century Immanuel Kant and Johann Lambert imagined the Universe as an infinite hierarchy of worlds. Later this type of Universe was also invoked to explain Olbert's paradox that the sky is dark at night, in contradiction with the assumption of an infinite number of stars. The mathematical description of fractals only came with Mandelbrot (Mandelbrot (1982)), that reinterpreted them with the concepts of self-similarity and scale-invariance. The denomination of fractal by Mandelbrot comes from the latin word *fractus* which means break. An exact definition of a fractal is difficult since it would exclude possible interesting cases. In Falconer (1990) a loose definition is given for a fractal *F*:

" *1. F has a fine structure, i.e. details on arbitrary small scales.*
 2. F is too irregular to be described in traditional geometrical language, both locally and globally.
 3. Often F has some form of self-similarity, perhaps approximate or statistical.
 4. Usually, the 'fractal dimension' of F (defined in some way) is greater than its topological dimension.
 5. In most cases of interest F is defined in a very simple way, perhaps recursively. "

Fractals can be either deterministic, i.e. constructed by an iterative process or recursion, or they can be random, i.e. generated by a stochastic process. Usually, only fractals constructed using an iterative process, rigorously present a self-similarity property. Random fractals are the most used in practice, and can describe numerous irregular objects in the real world, for example clouds, mountains, coastlines or turbulent fluids.

4.1.1 Fractal dimensions

Fractal dimension are extensions of traditional notions of dimensions, that exist for simple usual objects as lines, surfaces, and volumes. A first dimension is the box-counting dimension (also called Minkowski-Bouligand or Minkowski dimension). It is calculated by filling the space with small cubes of size δ, and counting the number of cubes that intersect the object (figure 6). For a line of length l, the number of such cubes $N(\delta)$ will be roughly $N(\delta) \approx \frac{l}{\delta}$, and for a surface of area A the number will be roughly $N(\delta) \approx \frac{A}{\delta^2}$.
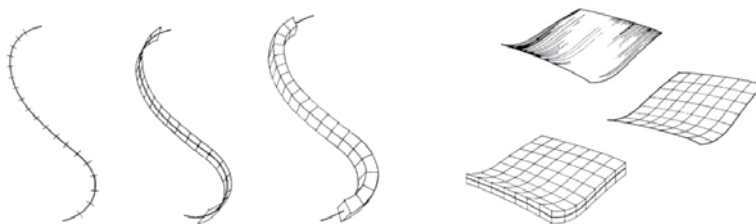

Fig. 6. Measuring the box-counting dimension of a line (left panel) and a surface (right panel). For a line the number of intersecting cubes of size δ scales as $\frac{1}{\delta}$, and for a surface it scales as $\frac{1}{\delta^2}$. Figure courtesy of Feder (1988).

The box-counting dimension of a set F is defined as

$$\text{dim}_{\text{box}}(F) = \lim_{\delta \to 0} \frac{\log(N(\delta))}{\log(1/\delta)} \tag{31}$$

If the limit does not exist, one must talk about the upper box dimension ($\text{dim}_{\text{upperbox}}$) and lower box dimensions ($\text{dim}_{\text{lowerbox}}$) as the upper and lower limits.

Another important dimension is the Hausdorff dimension. A r-cover of F is a countable collection of sets $\mathcal{A} = \bigcup_i \mathcal{A}_i$ which all have a diameter less than r, and such that F is totally covered by this collection. The d-dimensional Hausdorff measure of S is defined as

$$H^d(F) = \lim_{r \to 0} \inf_{\mathcal{A}} \sum_{i=0}^{\infty} diam(\mathcal{A}_i)^d \tag{32}$$

The Hausdorff dimension d_H is defined such that

$$H^d(F) \to \begin{cases} 0 \text{ for } d > d_H(F) \\ \infty \text{ for } d < d_H(F) \end{cases} \tag{33}$$

It can be shown that the Hausdorff dimension is always less than the box-counting dimensions

$$d_H(F) \leq \text{dim}_{\text{lowerbox}} \leq \text{dim}_{\text{upperbox}} \tag{34}$$

Another fractal dimension, which is well adapted to describe random point patterns, is the correlation dimension. In particular, it is very useful for measuring the fractality of the galaxy distribution. Noting $N(< r)$ the number of neighbors that a point has in average within a distance r, the point pattern is said to be fractal with the correlation dimension D_2 when

$$N(< r) \propto r^{D_2} \tag{35}$$

In other terms, D_2 is defined as

$$D_2 = \frac{d(\log N(< r))}{d(\log r)} \tag{36}$$

Such a measure is locally self-similar, i.e. the number of points varies a power law around every point. Intuitively, D_2 gives a local description of the measure linked to the point process. If $D_2 = 1$ then the points will be located on structures that are locally filamentary, and if $D_2 = 2$ these local structures will be planar.

For many well behaved fractals, all these dimensions are equal. For usual geometrical objects (lines, surfaces, volumes), they are integers, thus representing very particular cases. For purely fractal sets however, the Hausdorff dimension can be non integer. Let us give an example of such a case. Figure 7 shows the construction of the von Koch curve F, with segments of the curve each time subdivided into 4 new sub-segments. For this set, the fractal dimensions agree and are equal to $\frac{\ln 3}{\ln 2}$. Concerning the length of the curve E_k at iteration k, it is equal to $\left(\frac{4}{3}\right)^k$, and thus the limit curve has infinite length. On the other hand, the curve occupies zero area in the plane, so neither its length nor its area are useful descriptors.
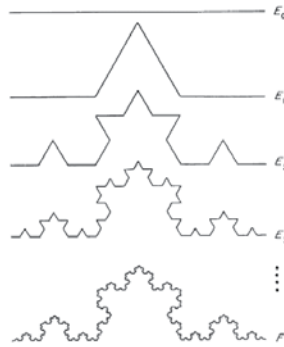
Fig. 7. Successive iterations in the construction of the von Koch curve F. Figure courtesy of Falconer (1990)

4.2 Multifractal measures

A multifractal measure μ on a set A is a generalization of a fractal. It is used when a single exponent (the fractal dimension) is not enough to describe the measure. It is everywhere self-similar, i.e. it varies locally as a power law at every point of its support.

4.2.0.1 Hölder exponent

The Hölder exponent or singularity exponent at a point x_0 is defined

$$\alpha(x_0) = \lim_{\delta \to 0} \frac{\log \mu(B_{x_o}(\delta))}{\log \delta} \tag{37}$$

where $B_{x_0}(\delta)$ is a ball centered on x_0 of radius δ. We have

$$\mu(B_{x_o}(\delta)) \propto \delta^{\alpha(x_0)} \tag{38}$$

The smaller the value of $\alpha(x_0)$, the less the measure is regular around x_0.

4.2.0.2 Singularity Spectrum

As previously mentioned, a multifractal measure is not characterized by a single exponent. For each value of α, a quantity of interest is the fractal dimension $f(\alpha)$ of the set of points that share this singularity exponent α

$$f(\alpha) = d_F \left(\{ x_0 \in A \,|\, \alpha(x_0) = \alpha \} \right) \tag{39}$$

The function $f(\alpha)$ is denoted as the singularity spectrum of μ. The singularity spectrum describes the fraction of points with a given exponent. So if we cover A with boxes of size δ, the number of boxes containing a point with exponent α will scale as $N_\alpha(\delta) \propto \delta^{-f(\alpha)}$, for small δ. The function $f(\alpha)$ is usually a single-humped function with its maximum at $\max_\alpha f(\alpha) = D$, where D is the dimension of the measure support. In the simple case of a monofractal, the function $f(\alpha)$ is reduced to a single point.

In order to calculate the singularity spectrum, the box-counting method consists in calculating the partition function

$$Z(q, \delta) = \sum_{i=0}^{N(\delta)} (\mu_i(\delta))^q \tag{40}$$

where $N(\delta)$ is the number of boxes of size δ intersecting the measure support, and we denoted $\mu_i(\delta) = \mu(B_i(\delta))$. If the measure μ is multifractal, $Z(q, \delta)$ follows a power law in the limit $\delta \to 0$ with

$$Z(q, \delta) \propto \delta^{\tau(q)} \tag{41}$$

The function $\tau(q)$ is related to the singularity spectrum by a Legendre transform

$$f(\alpha) = \min_q (q.\alpha - \tau(q)) \tag{42}$$

In the case of a monofractal, one has $\alpha = f(\alpha) = D$, with D the fractal dimension of μ (because the set of points with exponent α is the support of the measure itself). In terms of the Legendre transform this corresponds to $\tau(q) = D(q-1)$, i.e. the behavior of $\tau(q)$ versus q is a straight line with a slope given by the fractal dimension.

Some years ago, a new method based on wavelets, called Wavelet Transform Modulus Maxima (Muzy et al. (1994)), has shown to provide a more precise method for calculating the singularity spectrum.

4.2.0.3 Rényi dimension

The generalized fractal dimension also called Rényi dimension of order q is given by

$$D_q = \frac{\tau(q)}{q - 1} \tag{43}$$

D_0 corresponds to the Hausdorff dimension, D_1 is called the information dimension and D_2 is the correlation dimension, already defined in equation (36).

In practice, because of the limits in resolution and noise, the exponents can only be computed for a limited range of scales. Thus Rényi dimensions depend on the range of scales considered. In the example of the distribution of galaxies, the partition function can be written in a slightly different way

$$Z(q, r) = \frac{1}{N} \sum_{i=0}^{N} n_i(r)^{q-1} \tag{44}$$

where $n_i(r)$ is the numbers of neighbours of point labeled by i within a sphere of radius r. Scale-dependent Rényi dimensions $D_q(r)$ are defined as

$$D_q(r) = \frac{1}{q - 1}.\tau(q, r) = \frac{1}{q - 1} \frac{d \log Z(q, r)}{d \log r} \tag{45}$$

We show section 4.3.2 how these dimensions can be used in order to test the homogeneity of the galaxy distribution at different scales.

4.3 Fractality of large-scale structures

While it is clear that the distribution of galaxies is fairly inhomogeneous at small scales, the homogenization at larger scales has long been an debated question. This was due to

the lack of data to provide a definitive answer. For a detailed discussion on this matter at that period of time, we refer to Davis (1997). This is an important question, since the Friedmann-Robertson-Walker metric presumes large scale homogeneity and isotropy of the Universe (the cosmological principle). This is the simplest cosmological model and the cornerstone of modern cosmology, so it is a fair question to ask the degree to which it is supported by the observational evidence.

Due to the fractality over a wide range of scales, early redshift surveys did not show a trend to homogenization, and revealed always larger coherent structures (e.g. the CfA stickman, da Costa et al. (1994)). As we show section 4.3.1, the correlation function is not a well-defined quantity for a fractal distribution. Indeed it presupposes the existence of a mean density $\bar{\rho}$, which is not the case for a fractal distribution. So when estimating the correlation function on a fractal distribution, the expected result will have a systematic change with the sample size. In particular, the correlation length r_0 (defined such that $\xi(r_0) = 1$) increases linearly with the sample size. This systematic change in r_0 was observed in early galaxy surveys, which supported the fractality of the distribution. Another argument in favor of a fractal Universe was given by more luminous galaxies and clusters. They have been found to be more correlated than standard galaxies, and therefore to form larger coherent structures.

While all these arguments are valid, they are also explained in the standard model, that assumes an homogenization at large scale. The unstable behavior of r_0 is explained by the small scale fractality of the distribution, whereas the difference of clustering of different object populations is explained by the 'bias' of rare events, luminosity bias, and environmental effects. On the other hand, there are solid arguments in favor of an homogenization of the distribution. This phenomenon is referred as the End of Greatness, i.e. scales where no coherent structures can be observed anymore (thus one does not have to find new superlatives to describe structures).

Let us briefly present the numerous observations in favor of the standard picture, with a large-scale homogeneity of the matter distribution. The first type of observations is given by 2-dimensional data sets on the celestial sphere, including X-ray counts, radio sources counts, γ-ray bursts and CMB, which all present a remarkable degree of isotropy. If we do not assume our position to be special in the Universe, this 2D isotropy is a strong indication of homogeneity. Indeed, it is not clear at all, how random fractal distributions could have such isotropic 2D projections.

The second type of observations is given by 3-dimensional data, in particular the most recent redshift surveys, the 2dF and SDSS, which present a high degree of homogeneity at large scales. As an example of test, in Hogg et al. (2005) the mean number of points at distance R of LRG galaxies is studied and found to vary as $N(R) \propto R^3$ for scales $R > 70h^{-1}$Mpc. This is not compatible with a strictly fractal model, where this number would scale as $N(R) \propto R^D$, with $D < 3$ the fractal dimension. Another type of 3D data is given by the Lyman-α absorption lines, observed in the spectra of quasars, due to intervening clouds of neutral hydrogen. These clouds are found to be ubiquitous and nearly uniformly distributed in space.

4.3.1 Fractility and correlation function

There is a basic assumption behind the notion of correlation function, that the distribution has a well defined mean density $\bar{\rho}$. Indeed the correlation function measures the excess of probability to find a point at distance r of another given point, compared to an unclustered random distribution. The probability dP_r to find a galaxy in a volume dV which is separated by r of a given galaxy is

$$dP_r = \bar{\rho}(1 + \xi(r))dV \tag{46}$$

For a fractal field, the mean density $\bar{\rho}$ is not a well defined concept, thus the correlation function is also not well defined. As we show, if we still try to estimate the 'correlation function' of a fractal field, the expected result will depend on the volume size. Indeed, it can be proven that for a fractal of dimension D, the estimated 'correlation function' in a sphere of size R_s is given by (Sylos Labini et al. (1998))

$$\xi(r) = \frac{D}{3}\left(\frac{r}{R_s}\right)^{D-3} - 1 \tag{47}$$

This behavior is observed, but only at small scales, when measuring the correlation function on redshift surveys with $\xi(r) \propto r^{-\gamma}$ and $\gamma \approx 1.7$.

The correlation length r_0 is defined as the distance where $\xi(r_0) = 1$, i.e. the transition from strong to weak clustering. When estimating ξ on a fractal, the correlation length depends on the size of the volume (Pietronero (1987)). Using formula (47) we get

$$r_0 = \left(\frac{6}{D}\right)^{1/D-3} R_s \tag{48}$$

The correlation length is thus proportional to the volume size R_s. This behavior was observed in early galaxy surveys but more recent redshift surveys, from the Infrared Astronomical Satellite (IRAS, Neugebauer et al. (1984)) to the 2dF and SDSS, have contradicted this trend. The correlation length has become very stable for a given galaxy population ($r_0 \approx 5 - 6h^{-1}$ Mpc).

4.3.2 Multifractality and transition to homogeneity

The multifractal approach has been applied to the distribution of galaxies for this first time in Jones et al. (1988). In this study made on the CfA survey, the distribution of galaxies is found to be better described by a multifractal than by a monofractal. Thus, even for a small survey size, where the galaxy distribution is very inhomogeneous, it did not appear to be purely fractal.

More recent works have applied the multifractal formalism on the SDSS survey and SDSS mock catalogues. The method of analysis consists in studying the Rényi dimensions $D_q(r)$ as defined equation (45) for different scales r. If the exponents are found to be constant and equal to the euclidian dimension, this means that the galaxy distribution is homogeneous for the scales considered.

In Yadav et al. (2005) the exponents $D_q(r)$ are calculated on 2D slices of the SDSS Data Release 1 as well as ΛCDM mock galaxy catalogues, with different mass-luminosity bias (see section 2.3). Calculated in the range of scales from $60 - 70\,h^{-1}$Mpc to $150\,h^{-1}$Mpc, exponents $D_q(r)$ are found to vary between 1.7 and 2.2, i.e. close to a constant value equal to the dimension of the 2D slices. Furthermore, these results have been found to be consistent with the biased mock galaxy catalogues.

In Sarkar et al. (2009) the same analysis is performed on SDSS Data Release 6, but this time using the full 3D survey. This enables to reduce uncertainties in the estimation of the Rényi dimensions $D_q(r)$. Results at two different scales $r = 60\,h^{-1}$Mpc and $r = 70\,h^{-1}$Mpc are shown in figure 8. At scale $r = 60\,h^{-1}$Mpc, Rényi dimensions show small deviations from the constant value $D = 3$ equal to the dimension of the space. They are not consistent with

random catalogues (i.e. with randomly located points, homogeneous by construction) at the 1σ level. However at scale $r = 70\,h^{-1}\mathrm{Mpc}$, the Rényi dimensions are nearly constant with q and close to the value $D = 3$. They also become consistent with results obtained on random catalogues. The conclusion of these studies is that the galaxy distribution is homogeneous at large scale, with a transition to homogeneity at around $70\,h^{-1}\mathrm{Mpc}$.
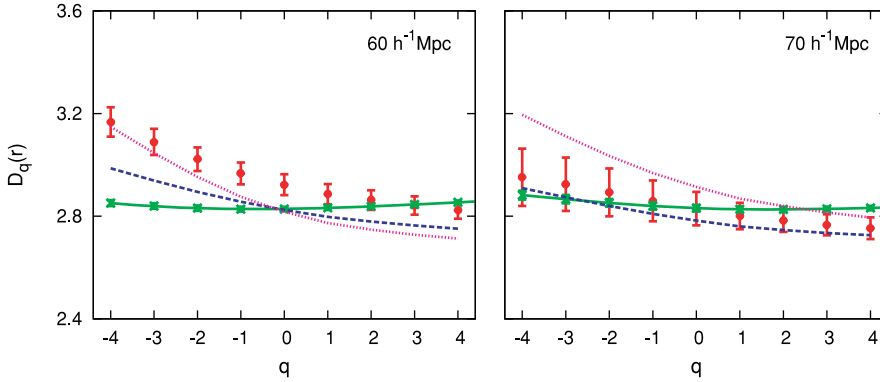

Fig. 8. The Rényi dimensions $D_q(r)$ as a function of q for $r = 60\,h^{-1}\mathrm{Mpc}$ (left panel) and $r = 70\,h^{-1}\mathrm{Mpc}$ (right panel). The figure shows the SDSS data (red points), random data (green solid line), N-body (blue dashed) and Millennium N-body simulation (pink dotted). Five independent N-body realizations are used to estimate the 1σ error bars shown on the SDSS data. Figure courtesy of Sarkar et al. (2009)

We conclude this section by quoting Davis (1997), which sums up the current accepted view of the galaxy distribution having a fractal behavior at small scale, and at the same time getting homogeneous at large scales:
"The observed galaxy distribution, being a real physical system rather than a mathematical idealization, is a beautiful example of a limited-scale fractal joined onto sensible, dynamically evolving outer boundary conditions."

5. Conclusion

We have presented several methods for studying the galaxy distribution, which are sensitive to different aspects of it. They should not be directly compared, but rather be considered as complementary. Their goal is to test cosmological models, and constrain parameters inside them. Ideally, the statistics are given by analytic formulas for each model, which can be compared to data results. N-body simulations can also be used for this task, but one has to trust these simulations, which are usually done for a single set of cosmological parameters. The most popular method is Fourier analysis, based on the second order statistics, which are the correlation function and power spectrum. One advantage is that they can be analytically predicted at linear scales, for usual ΛCDM and wCDM models. Another advantage is that BAOs have their signature directly imprinted in these statistics, as they manifest by an excess of clustering at the sound horizon scale. BAOs in the correlation function (or power spectrum) confirms the usual cosmological paradigm of linear gravitational evolution from redshift $z \approx 1000$, and the existence of dark energy. They also provide a standard ruler to quantify the evolution of distances, i.e. constrain cosmological parameters.

A second type of methods studies the global morphology and topology of the galaxy field, through Minkowski functionals. They can also be analytically predicted at linear scales where the matter distribution is gaussian. They were first used to determine the transition from linear to non-linear regimes, where the field becomes non gaussian. Recently they have been used as a standard ruler to constrain cosmological parameters, since their value per unit volume is predicted analytically. They have also been used to distinguish between models with different morphologies, that are well differentiated by Minkowski functionals.

Finally we presented the fractal analysis of the galaxy distribution, also probing the global morphology of the field. Fractality of the galaxy distribution is well-established at small scales. On the other hand, observations confirm large-scale homogeneity, as required by the cosmological principle which is the cornerstone of modern cosmology. One has to turn to multifractal analysis to model the galaxy distribution and study this transition to homogeneity. This is done by studying the scaling exponents of the distribution, which are found to approach the dimension of the space at scales around $70h^{-1}$Mpc, indicating large-scale homogeneity.

6. References

Bassett, B. & Hlozek, R. (2010). *Baryon acoustic oscillations*, Cambridge, UK.

Blake, C. et al. (2011). The WiggleZ Dark Energy Survey: testing the cosmological model with baryon acoustic oscillations at z=0.6, *ArXiv e-prints* .

Bond, J. R., Kofman, L. & Pogosyan, D. (1996). How filaments of galaxies are woven into the cosmic web, *Nature* 380: 603–606.

Cabré, A. & Gaztañaga, E. (2011). Have baryonic acoustic oscillations in the galaxy distribution really been measured ?, *Monthly Notices of the Royal Astronomical Society* 412: L98–L102.

Colless, M. et al. (2001). The 2dF Galaxy Redshift Survey: spectra and redshifts, *Monthly Notices of the Royal Astronomical Society* 328: 1039–1063.

Cooray, A. & Sheth, R. (2002). Halo models of large scale structure, *Physics Reports* 372: 1–129.

da Costa, L. N. et al. (1994). A complete southern sky redshift survey, *Astrophysical Journal Letters* 424: L1–L4.

Davis, M. (1997). Is the Universe Homogeneous on Large Scales?, *Critical Dialogues in Cosmology*, pp. 13–+.

Davis, M. & Peebles, P. J. E. (1983). A survey of galaxy redshifts. V – The two-point position and velocity correlations, *Astrophysical Journal* 267: 465–482.

Eisenstein, D. J. & Hu, W. (1998). Baryonic Features in the Matter Transfer Function, *Astrophysical Journal* 496(2): 605.

Eisenstein, D. J. et al. (2005). Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies, *Astrophysical Journal* 633: 560–574.

Falconer, K. (1990). *Fractal Geometry*, Mathematical Foundations and Applications, John Wiley & Sons Ltd., New-York.

Feder, J. (1988). *Fractals*, Plenum Press, New York.

Gott, J., Dickinson, M. & Melott, A. (1986). The sponge-like topology of large-scale structure in the universe, *Astrophysical Journal* 306: 341–357.

Hamilton, A. J. S. (1993). Toward better ways to measure the galaxy correlation function, *Astrophysical Journal* 417: 19.

Hikage, C. et al. (2003). Minkowski Functionals of SDSS Galaxies I : Analysis of Excursion Sets, *Publications of the Astronomical Society of Japan* 55: 911–931.

Hogg, D. W. et al. (2005). Cosmic homogeneity demonstrated with luminous red galaxies, *Astrophysical Journal* 624: 54–58.

Hütsi, G. (2006). Acoustic oscillations in the SDSS DR4 luminous red galaxy sample power spectrum, *Astronomy and Astrophysics* 449: 891–902.

Jones, B. J. T., Martinez, V. J., Saar, E. & Einasto, J. (1988). Multifractal description of the large-scale structure of the universe, *Astrophysical Journal Letters* 332: L1–L5.

Kaiser, N. (1986). A sparse-sampling strategy for the estimation of large-scale clustering from redshift surveys, *Monthly Notices of the Royal Astronomical Society* 219: 785–790.

Kerscher, M., Szapudi, I. & Szalay, A. S. (2000). A Comparison of Estimators for the Two-Point Correlation Function, *Astrophysical Journal* 535: L13–L16.

Labatie, A., Starck, J., Lachièze-Rey, M. & Arnalte-Mur, P. (2010). Uncertainty in 2-point correlation function estimators and BAO detection in SDSS DR7, *ArXiv e-prints* .

Landy, S. D. & Szalay, A. S. (1993). Bias and variance of angular correlation functions, *Astrophysical Journal* 412: 64–71.

Mandelbrot, B. (1982). *The Fractal Geometry of Nature*, Times Books.

Martínez, V., Starck, J., Saar, E., Donoho, D., Reynolds, S., de la Cruz, P. & S., P. (2005). Morphology of the Galaxy Distribution from Wavelet Denoising, *Astrophysical Journal* 634: 744–755.

Muzy, J. F., Bacry, E. & Arnéodo, A. (1994). The multifractal formalism revisited with wavelets, *International Journal of Bifurcation and Chaos* 4: 245–302.

Neugebauer, G. et al. (1984). The Infrared Astronomical Satellite (IRAS) mission, *Astrophysical Journal Letters* 278: L1–L6.

Norberg, P. & Baugh, C. M. (2001). The 2df galaxy redshift survey: The dependence of galaxy clustering on luminosity and spectral type, *Monthly Notices of the Royal Astronomical Society* 332: 827. 13 p.

Park, C. & Kim, Y. (2010). Large-scale Structure of the Universe as a Cosmic Standard Ruler, *Astrophysical Journal Letters* 715: L185–L188.

Peebles, P. J. E. & Hauser, M. G. (1974). Statistical Analysis of Catalogs of Extragalactic Objects. III. The Shane-Wirtanen and Zwicky Catalogs, *Astrophysical Journal Supplement Series* 28: 19–+.

Percival, W. J. et al. (2010). Baryon acoustic oscillations in the Sloan Digital Sky Survey Data Release 7 galaxy sample, *Monthly Notices of the Royal Astronomical Society* 401: 2148–2168.

Pietronero, L. (1987). The fractal structure of the universe: Correlations of galaxies and clusters and the average mass density, *Physica A Statistical Mechanics and its Applications* 144: 257–284.

Pons-Bordería, M., Martínez, V. J., Stoyan, D., Stoyan, H. & Saar, E. (1999). Comparing estimators of the galaxy correlation function, *Astrophysical Journal* 523: 480–491.

Sarkar, P., Yadav, J., Pandey, B. & Bharadwaj, S. (2009). The scale of homogeneity of the galaxy distribution in SDSS DR6, *Monthly Notices of the Royal Astronomical Society* 399: L128–L131.

Sheth, J. V., Sahni, V., Shandarin, S. F. & Sathyaprakash, B. S. (2003). Measuring the geometry and topology of large-scale structure using SURFGEN: methodology and preliminary results, *Monthly Notices of the Royal Astronomical Society* 343: 22–46.

Springel et al. (2005). Simulations of the formation, evolution and clustering of galaxies and quasars, *Nature* 435: 629–636.

Sylos Labini, F., Montuori, M. & Pietronero, L. (1998). Scale-invariance of galaxy clustering, *Physics Reports* 293: 61–226.

Tomita, H. (1990). *Formation, Dynamics and Statistics of Patterns, Vol.1, (World Scientific), 113*, Vol. 1, World Scientific, pp. 113–157.

Yadav, J., Bharadwaj, S., Pandey, B. & Seshadri, T. R. (2005). Testing homogeneity on large scales in the Sloan Digital Sky Survey Data Release One, *Monthly Notices of the Royal Astronomical Society* 364: 601–606.

York, D. G. et al. (2000). The Sloan Digital Sky Survey: Technical Summary, *Astrophysical Journal* 120: 1579–1587.

Zehavi, I. et al. (2005). The Luminosity and Color Dependence of the Galaxy Correlation Function, *Astrophysical Journal* 630: 1–27.

Zunckel, C., Gott, J. R. & Lunnan, R. (2011). Using the topology of large-scale structure to constrain dark energy, *Monthly Notices of the Royal Astronomical Society* 412: 1401–1408.

Edited by Adnan Ghribi

The twentieth century elevated our understanding of the Universe from its early stages to what it is today and what is to become of it. Cosmology is the weapon that utilizes all the scientific tools that we have created to feel less lost in the immensity of our Universe. The standard model is the theory that explains the best what we observe. Even with all the successes that this theory had, two main questions are still to be answered: What is the nature of dark matter and dark energy? This book attempts to understand these questions while giving some of the most promising advances in modern cosmology.

IntechOpen