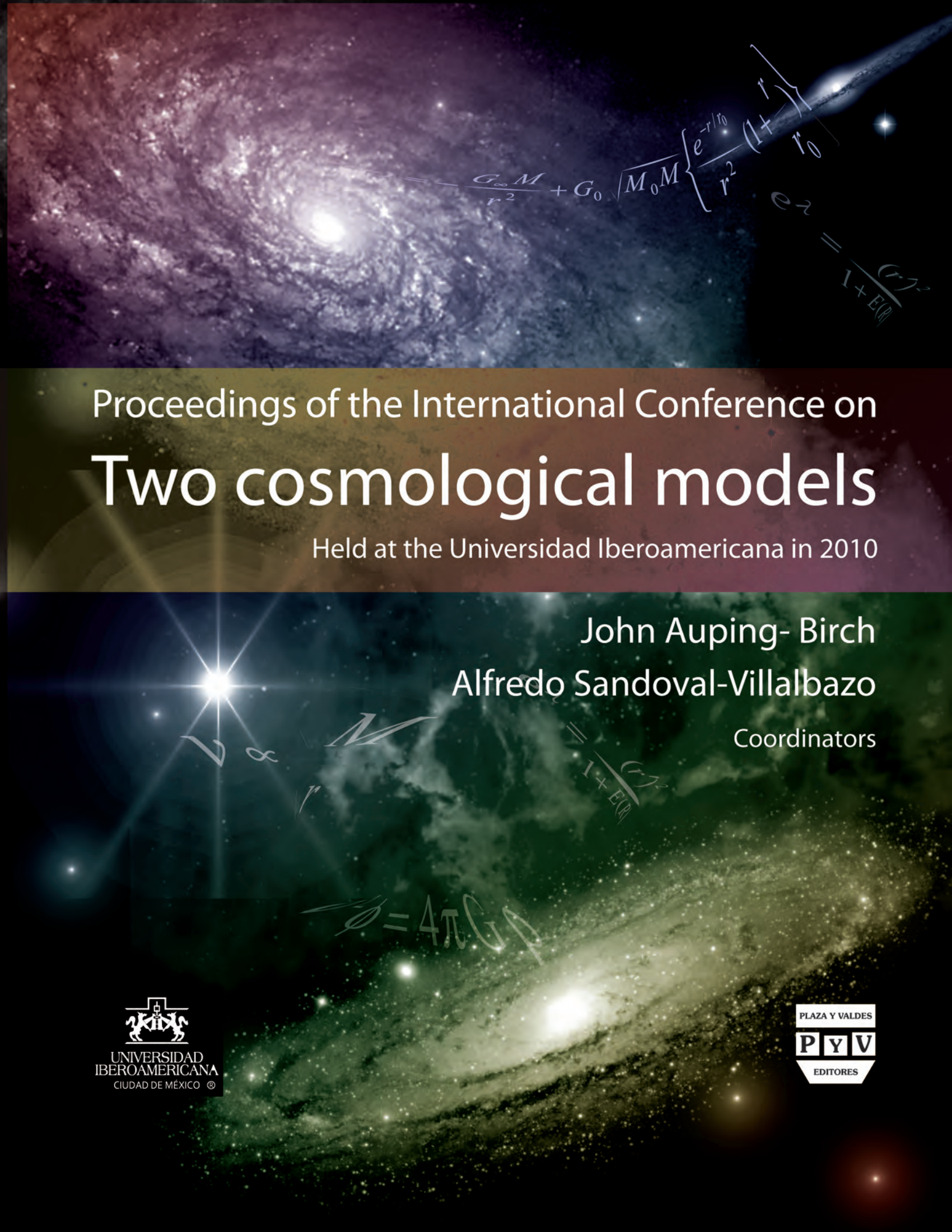


Proceedings of the International Conference on

Two cosmological models

Held at the Universidad Iberoamericana in 2010

John Auping- Birch
Alfredo Sandoval-Villalbazo
Coordinators

UNIVERSIDAD IBEROAMERICANA
CIUDAD DE MÉXICO ®

PLAZA Y VALDES
PYV
EDITORES

Proceedings of the International Conference on
Two Cosmological Models

PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON

TWO COSMOLOGICAL MODELS

HELD AT THE UNIVERSIDAD IBEROAMERICANA IN 2010

MEXICO

2012

John A. Auping

Coordinator

UNIVERSIDAD IBEROAMERICANA
CIUDAD DE MÉXICO ®

PLAZA Y VALDES
PYV
EDITORES

Primera edición: 2012

Universidad Iberoamaericana
Prolongación Paseo de la Reforma 880
Lomas de Santa Fe, Mexico, D.F.

John A. Auping
jauping@iwm.com.mx

Plaza y Valdés, S. A. de C. V.

Plaza y Valdés, S.A. de C.V.
Manuel María Contreras 73. Colonia San Rafael
México, D.F. 06470. Teléfono: 50 97 20 70
editorial@plazayvaldes.com
www.plazayvaldes.com

Plaza y Valdés Editores
Calle Murcia, 2. Colonia de los Ángeles
Pozuelo de Alarcón 28223, Madrid, España
Teléfono: 91 862 52 89
madrid@plazayvaldes.com
www.plazayvaldes.es

ISBN: 978-607-402-530-9

Impreso en México / Printed in Mexico

Contents*

*The pages have double numeration. The number on top is the book page number and the one below is the lecture page number.

[1] Lectures given on November 17th , 2010
[2] This lecture was not made available for its publication in the Proceedings
[3] Lectures given on November 18th , 2010
[4] This lecture was not made available in its written form, so that a transcription of its oral presentation is published
[5] This lecture was not made available for its publication in the Proceedings

PART III, THE APPARENT ACCELERATION OF THE EXPANSION OF THE UNIVERSE[7]

[6] Hans Ohanian's comments on Philip Mannheim's lecture, made available after the Conference. Philip Mannheim was invited to express his comments on these comments but preferred not to do so

[7] Lectures given on November 19th , 2010

Preface

The International Conference on Two Cosmological Models was held at Universidad Iberoamericana in Mexico City, Mexico, from November 17th to 19th, 2010, as a forum devoted to the study and discussion of two important problems of modern cosmology. Of the more than 50 experts of the traditional ΛCDM model that were invited by the University to assist at the Conference, five accepted our invitation. Of the 11 representatives of alternative models that were invited, all agreed to participate. Two participants gave an overview of historical and relativistic aspects of the problem, without aligning with any model in particular. A total of 18 speakers from Brazil, Canada, Chile, France, Germany, India, Mexico, the Netherlands, New Zealand, and the USA gave a total of 20 lectures and participated in an open discussion on the following two topics: 1) The concept of dark matter as a possible explanation of the rotation velocity of galaxies and galaxy clusters in the context of Newtonian dynamics; and the alternative explanation through Einstein's general relativity, without dark matter. 2) The concept of dark energy as a possible explanation of the apparent acceleration of the expansion of the universe; and the alternative explanation through Einstein's gravitational theory, without dark energy.

The overall impact of the event was more than satisfactory. These Proceedings contain the lectures on the topics covered in the International Conference on Two Cosmological Models, except for two of them, who could not send us the written version of their lecture. One participant did not send us the written version of his lecture, but gave us permission to transcribe its verbal version. After the Conference, some participants made their lectures available in ArXiv, adding some references to more recent essays, published after the Conference, so that these Proceedings provide the reader with an update of the most current research on these very transcendental topics.

We would like to thank everyone who contributed to the success of the International Conference on Two Cosmological Models. Very special thanks are due to the invited speakers who addressed a very interesting and high quality set of talks and shared their deep knowledge and time with the participants.

We grately acknowledge Dr. John Auping-Birch and all the staff of Universidad Iberoamericana for the warm hospitality, which was extended to all the participants. We specially thank Dr. José Morales-Orozco, Rector of Universidad Iberoamericana, Mexico City for sponsoring this international endeavor. We hope that these Proceedings will serve to foster the impressive growth of high precision cosmology and the discussion on the different theoretical interpretations of its findings and, additionally, reinforce the existing ties between the Mexican researchers and scientists from all over the world.

Dr. Alfredo Sandoval Villalbazo Director of the Physics and Mathematics Department Universidad Iberoamericana Mexico, July 2012

PART I

SPIRAL GALAXY ROTATIONAL VELOCITY

GRAVITATION AND SPACETIME: EINSTEIN'S CONTRIBUTION

Hans C. Ohanian
Department of Physics, University of Vermont, Burlington, VT 05405-0125, USA
hohanian@uvm.edu

Abstract

As I argued in Einstein's Mistakes [1], most of Einstein's great discoveries rest on conceptual mistakes which he used as stepping stones toward a final, true, result. This lecture is a summary of the various mistakes that paved Einstein's progress toward his theories of special and general relativity. It also includes a detailed discussion of an extra, previously unrecognized, mistake in the application of the equivalence principle, namely, that the gravitational redshift cannot be derived from this principle by Einstein's 1911 argument, because the equivalence principle contains a contradiction that renders it invalid when applied in flat spacetime.

1 Introduction

I want to begin this lecture by making a wish: I earnestly and passionately wish that the participants of this conference will make very many big mistakes. My wish is not malicious, because I believe it is by making great mistakes that we make great discoveries. As James Joyce said, "Errors are the portals of discovery." And I am going to illustrate this maxim by showing you how Einstein's great and wonderful mistakes led him to his great and wonderful discoveries in special and general relativity.

Although the modern view of space and time was not exclusively the work of Einstein, he made the most fundamental and most profound contributions, and for many years he was the dominant figure in relativistic physics, as well as the dominant figure in all of physics. Einstein became a celebrity, adored by the public, and even today, fifty five years after his death, his celebrity status survives. If you google Einstein you get 155 million hits, which is only slightly below the number for Jesus.

Bernard Shaw compared Einstein to the great conquerors in history and called him a maker of universes. He said "Ptolemy made a universe, which lasted 1400 years. Newton, also, made a universe, which lasted 300 years. Einstein has made a universe, and I can't tell you how long it will last" [2] .We are now in the 105^{th} year of Einstein's spacetime universe, and so far all is well. I don't regard the theories that will be presented at this conference as an attempt to overthrow Einstein's view of spacetime—these theories merely adjust and refine Einstein's work.

The only serious attempt at overthrowing Einstein's spacetime is found in string theory, but so far, this has been a chaotic endeavor, without any sharply defined target or any clearcut success. String theorists claim they will someday be able to predict everything, but so far they have predicted next to nothing. Last week, in Physics Today, I finally read a prediction made by a string theorist who had investigated neutrinos, and who announced "We showed that in no case could the theory generate light but not massless neutrinos. That work presents a clear

example of a test of string theory" [3] . What this sentence means is a mystery to me; I have read it forward and backward and sideways, and I still am not sure what it means. As best I can tell, this string theorist seems to announce that his theory can generate only neutrinos of mass zero, which, of course, disagrees with the observational fact that neutrinos have mass. And he regards this as progress?

Because Einstein played such a preeminent role in relativity, every development in relativity has tended to be credited to him. The physicist and historian Jagdish Mehra thought this was a result of "the sociology of science, the question of the cat and the cream. Einstein was the big cat of relativity, and the whole saucer of its cream belonged to him by right and by legend, or so most people assume!" [4].

But legends are often wrong. Not all of the contributions to relativity came from Einstein, and even those contributions that originated from Einstein were often in need of corrections, improvements, and emendations. In this lecture, I will dissect the mistakes that Einstein made in the seminal papers that led to the development of special relativity and general relativity in 1905 and 1911-1916, and I will show how his great and surprising mistakes led to his great and surprising discoveries.

2 Special relativity

The groundwork for special relativity had already been laid by Einstein's predecessors, especially Hendrik Lorentz and Henri Poincare. Einstein did not admit to that in his first paper on relativity, but he admitted it later, saying that by 1905 relativity "was in the air" [5] . Einstein's 1905 paper contains several fundamental contributions. Not all of them were new; his statement of the principle of relativity had been anticipated by Poincare and his coordinate transformations were a rederivation of the Lorentz transformations obtained by Lorentz a year earlier, which Einstein had not noticed. And some of the contributions in the 1905 paper involved serious mistakes. Despite, and perhaps because of, these mistakes, this paper launched the theory of relativity by laying down a general program for how to implement the principle of relativity for all laws of physics by means of the Lorentz transformations.

Einstein begins his 1905 paper with a discussion of synchronization of clocks. This discussion is phrased in deceptively simple language, although it deals with a profound physical and philosophical question. He asks, What does it mean to say "that a train arrives here at 7 o'clock?" And he answers that it means that "the pointing of the small hand of his watch to 7 and the arrival of the train are simultaneous events" [6] . That much is simple. But things get more complicated when we want to synchronize clocks or events at different locations.To achieve synchronization at distant locations, say, Ciudad de México and Mérida, we need some special synchronization procedure, and Einstein decided to adopt a procedure of sending light signals back and forth between the two locations. If the light signal leaves here at noon, and comes back in 6 milliseconds, then it must have reached Mérida at noon plus 3 milliseconds, which tells our colleagues at the Universidad de Yucatán how to synchronize their clock with ours.

Einstein became obsessed with this synchronization procedure, and he believed it was the solution to the puzzle of the invariance of the speed of light, a problem he had wondered about since his teenage years. Due to the emphasis he gave to this procedure in his 1905 paper, it became known as the "Einstein procedure" for synchronization. But it did not originate with Einstein. It actually was a procedure that had been adopted in the 1850s, when the first long-distance telegraph lines were laid in the US, and astronomers decided to use back-and-forth telegraph signals to synchronize distant clocks, for use in accurate determinations of geographical longitude. Throughout the second half of the 19^{th} century, this method was widely used for transcontinental

geographical surveys, such as the surveys of India and of Russia; and, when the first transoceanic telegraph lines were laid, it was also used for intercontinental longitude determinations. In the first years of the 20^{th} century, the method was used for a new longitude determination of the Paris Observatory relative to Greenwich, and maybe that was where Einstein heard about this method and decided to imitate it.

Einstein argues that if we adopt this synchronization procedure for the clocks at the starting and ending points of a racetrack, it becomes meaningless to measure the separate one-way speeds of light forward and backward along the racetrack, because such measurements would be logically circular. Einstein therefore claims that the equality of one-way speeds of light is a stipulation, beyond the reach of actual experiment.

His logic is impeccable, but his physics is atrocious. Einstein forgets that there might be other ways to synchronize clocks, such as transport of clocks from one place to the other. Before telegraphy became available, such clock transport was widely used by astronomers in the 19 century for determinations of geographical longitude. To synchronize clocks at different stations, astronomers transported chronometers between the stations, often using several chronometers, to improve the accuracy by averaging. For instance, to determine the longitude difference between Greenwich and Valentia (on the west coast of Ireland), George Airy, then the Astronomer Royal, transported 30 chronometers from Greenwich to Valentia, and then he repeated this times, for an even better average [7].

And Einstein forgets that there might be other ways to measure the one-way speed of light without synchronization of clocks. As far back as the17^{th} and 18^{th} centuries, the Danish astronomer Olaf Roemer and the English astronomer James Bradley measured the one-way speed of light without making use of any clock synchronization. Roemer (1676) exploited the time delay in the observed eclipses of the satellites of Jupiter, and Bradley (1727) exploited the aberration of starlight. Both methods hinge, in essence, on comparing the speed of light with the speed of the Earth in its orbit.

Why or how Einstein overlooked these two well-known historical determinations of the speed of light is a puzzle. Roemer's and Bradley's determinations of the speed of light were well known in the 19^{th} century, and they were even discussed in introductory physics textbooks [8]. Maybe Einstein decided to dismiss these astronomical methods out of hand because, in practice, they could not achieve the high precision attained by terrestrial methods, first by Armand Fizeau in 1849 and then by Abraham Michelson in a series of measurements in the 1880s and 1890s. Michelson continued to improve these measurements, and for his work he came to be called the "master of light" (in Spanish, "el señor de la luz," which rather sounds like a religious title). But in his paper Einstein was discussing questions of principle, not questions of practice, and the practical limits of the precision of synchronization were not an issue.

Maybe Einstein deleted Roemer and Bradley from his mind because they were an inconvenient truth, awkward to fit into his own way of thinking about the speed of light and the synchronization problem. During the months of intense, feverish thinking that preceded the writing of the relativity paper Einstein was in the grip of a mystical, intuitive, and irrational inspiration. The historian Peter Galison described how Einstein, during an afternoon walk in the hills around Berne, suddenly arrived at the idea that the solution of the problem of a universal speed of light rested in the synchronization [9]. And the biographer Abraham Pais thought that perhaps this intense inspirational experience "was so overwhelming that it seared his mind and partially blocked out reflections and information that had been with him earlier" [10].

In the absence of Roemer and Bradley, Einstein drew the mistaken conclusion that the one-way speed of light must be established by stipulation, that is, by fiddling with the synchronization of clocks in such a way as to produce equal one-way speeds in opposite directions. Instead of promulgating the universal, constant value of the one-way speed of light as a stipulation, or a

dogma, Einstein should have treated it as simple postulate, that is, a proposed law or principle of physics to be tested and confirmed by experiment, which is what he did, quite correctly with the round-trip speed of light.

Einstein's stipulation was a mistake, but it was a fruitful mistake, because it permitted him to rush forward with the development of relativity and bypass the troublesome question of how to obtain conclusive experimental confirmation of the universal value of the one-way speed of light.

Here we see a crucial difference between Newton's and Einstein's approach to physics. Newton declared "I make no hypotheses." Whether that is quite true of all his work is debatable, but at least such was his announced intention. In contrast, Einstein made hypotheses whenever he could get away with it, and sometimes he even made hypotheses that were preempted and or even contradicted by known experimental and observational facts. He was a dogmatist, and he had the arrogance and stubbornness that goes with that. We see an example of this in his stipulation about the one-way speed of light and his cavalier rejection of Roemer's and Bradley's methods. We see another example in Einstein's hypothesis of quanta of light, which he based on Wien's law for the blackbody spectrum, in willful and deliberate defiance of Planck's law. By 1905, Planck's law was experimentally well established and known to be correct, and Einstein must have been aware that his hypothesis of thermal radiation as a gas of quanta, treated by classical statistics, was in conflict with Planck's law. He never mentioned this inconvenient truth, and he insisted on a hypothesis that—until the introduction of quantum statistics twenty years later—was in contradiction with the experimental evidence. And we see the most glaring example of Einstein's dogmatism in his rejection of the probabilistic interpretation of quantum mechanics and his famous declaration that "God does not play dice."

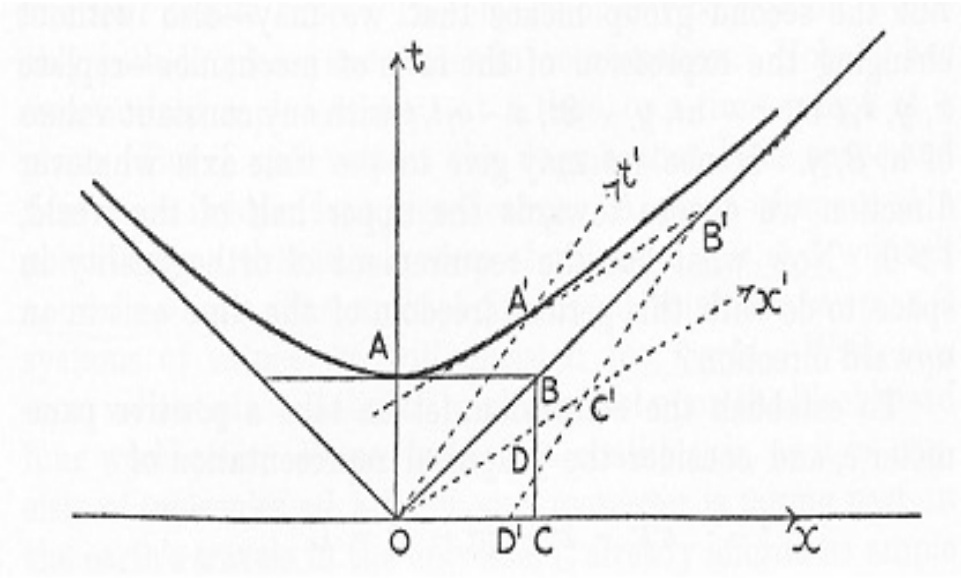
Newton believed, or pretended to believe, in a "bottom-up" approach to physics, based on experimental and observational facts from which, by a process of generalization, or induction, the laws of physics could be extracted. Einstein believed in a "top-down" approach to physics, by inspirational formulation of hypotheses from which consequences could be derived. As John Auping points out [11], Einstein's method is deductive, not inductive, and this is clearly revealed in a newspaper article Einstein wrote in 1919:

The truly great advances in our understanding of nature originated in a way almost diametrically opposed to induction. The intuitive grasp of the essentials of a large ensemble of facts leads the researcher to the formulation of one or more hypothetical fundamental laws. From the fundamental law (system of axioms) he draws his conclusions as completely as possible in a purely logical-deductive manner [12].

Einstein believed he had the intuitive insight to perceive the truth, and he, and only he, knew what hypotheses to make. He had the confidence of a visionary and of a fanatic, and he did not hesitate to go against experimental and observational facts when these did not fit his theories.

In the second section of his paper, Einstein exploits the constant one-way speed of light as the basis for a derivation of the transformation of space and time coordinates between two inertial reference frames. His treatment of this derivation is correct, but astoundingly clumsy. Even in abbreviated form presented in the printed publication, the calculations go on and on for five and a half tedious pages; and, as shown in the recent book on Kinematics by Alberto Martínez [13], if he had spelled out all the details of his calculations, it would have taken him thirty pages of print. The trouble was that Einstein failed to see that there is a simple physical argument to establish that lengths transverse to the velocity will not expand or contract: if two meter sticks in relative motion are oriented transversely, then, by symmetry of the physical situation, a contraction of the second relative to the first requires an equal contraction of the first relative to the second, which is contradictory, and rules out any contraction (or expansion) [14]. Today, this argument is familiar to all students of relativity. If the transverse lengths are unchanged,

Figure 1: Minkowsky's derivation of the Lorentz transformation



they can be ignored, and it then suffices to examine the transformations of the time coordinate and the longitudinal length coordinate, which makes the calculation much simpler.

How much simpler can be seen by taking a look at the derivation of the transformation equations presented by Hermann Minkowski in a lecture to the Congress of German Natural Scientists and Physicians in 1908. Minkowski had been one of Einstein's math professors at the Polytechnic in Zurich, and he had a rather poor opinion of Einstein—he called him a "lazy dog." When Einstein published his famous five papers in 1905, Minkowski was astounded; he said "I really would not have believed him capable of it" [15]. Minkowski then took the invariance of the speed of light from Einstein's relativity paper and produced a new, graphical derivation of the transformation equations. It consists of one simple diagram (Fig. 1), from which, in half a minute, you can deduce the transformation equations. In fact the derivation from this diagram is so simple that Minkowski did not even bother to spell it out in the published version of his lecture. I imagine that at the Congress he merely drew this diagram on a blackboard and said one or two dozen words about it. If you want to try to derive the transformation equations, here are the two dozen words you need to know: First draw the line OA' along the t' axis, then the line A'B' tangent to the hyperbola $t^2 - x^2 = 1$ at A'. The rest of the diagram is self-evident.

The difference between Einstein's five and a half messy pages of calculations and this simple diagram reveals the difference between a master mathematician such as Minkowski and his pupil. But, as Einsteins achievements show, you don't need to be a brilliant mathematician to become a great physicist. In fact, Einstein had little mathematical talent, and he had no real interest in mathematics. He thought in images, not in words or formulas. In an interview with a psychologist, he said: "I rarely think in words at all. A thought comes, and I may express it in words afterwards" [16]. In his autobiography, he excused his mathematical deficiencies saying, "I saw that mathematics was split up into many specialties, each of which could absorb the short lifespan granted to us.Thus I saw myself in the position of Buridan's ass, which was unable to decide on a particular bundle of hay" [17]. What he does not mention in this autobiography is that throughout his life he relied heavily on friends and on assistants to do his mathematics for

him. At the Polytechnic and for his early investigations of general relativity, he relied on his friend Marcel Grossman, who taught him the basics of Riemannian geometry. Later, when his career blossomed with his appointment in Berlin, he hired a personal mathematical assistant, and in later years he always had an assistant available; these came and went, a total of ten or so altogether, and they did all the calculations that Einstein found too tedious.

Besides improving on Einstein's mathematical presentation, Minkowski made a crucial contribution to our understanding of spacetime by recognizing that Einstein's theory of relativity implies a complete unification of space and time. The Lorentz transformation equations show that what is time and what is space depends on the reference frame, and that transformations of reference frame mix space and time. As Minkowski expressed it, "space by itself, and time by itself are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality" [18] .

This unification of space and time also caught the attention of philosophers, and of the public, and of poets and writers who delighted to insert esoteric references to unified space and time in their writings. Carlos Fuentes included a well-tuned phrase of this kind in one of his books: "Habrá solo la unidad total, olvidada, sin nombre y sin hombre que la nombre: fundidos espacio y tiempo, materia y energía" [19] (I can try to translate this, but it doesn't quite have the same ring in English, and it loses the charming musicality that Fuentes gave it: "There will be only total unity, forgotten, without a name and without anyone to give it a name: space and time are merged, and so are matter and energy") .

Minkowski deserves credit for his perception of the unification of space and time as well as for the unification of electric and magnetic fields, for which he invented the 16-component electromagnetic field tensor. But we must not forget that Einstein laid the groundwork for that, and without Einstein's 1905 paper, Minkowski would never have arrived at a unification of space and time. Half in serious, half in jest, Minkowsky also contributed a zany formula relating the units of space and time, $3 \times 10^5\ km = \sqrt{-1}\ sec$. This incorporates the speed of light, 300000 *km per sec*, and it incorporates the notion that, in a purely formal way, time can be treated as a mathematically imaginary fourth dimension of space. Minkowski called this his mystic formula. He would undoubtedly have made many more valuable contributions to special and to general relativity, but he died prematurely, in 1909.

Einstein was not pleased with Minkowski's mathematical approach. He called Minkowski's tensor formalism "unnecessary erudition," and said that "ever since the mathematicians have thrown themselves on the theory of relativity I can't understand it any more" [20]. At first he remained stubbornly opposed to the use of tensor formalism in the treatment of relativity, and he did not adopt the use of tensors until five or six years later, when he began to work on general relativity.

From his transformation formulas for the coordinates of inertial reference frames, Einstein deduced the time dilation and the length contraction: a moving clock runs slow, and a moving body, such as a meter stick, is contracted. He suggested that this implies that a clock on the equator of the Earth, moving with the Earth's rotation, will tick slower than a clock at the pole. This was an unfortunate choice of example, because the gravitational time dilation of the clock at the pole, which Einstein was to discover a few years later, actually compensates for the time dilation of the clock located at the equator. But Einstein must be given credit for taking the time dilation of moving clocks seriously, whereas Lorentz, who had deduced the same transformation laws a year earlier, attached no real physical significance to the change of clock rates implied by his transformation equations.

In his deduction of time dilation and length contraction Einstein commits a serious sin of omission: he never gives us any physical explanation of time dilation and length contraction. He treats these as purely abstract, mathematical consequences of the Lorentz transformations. The

latter, in turn, are consequences of the requirement of a constant, universal speed of light in all reference frames. This creates the impression of a teleological approach to physics: Are we to believe that clocks run slow and meter sticks contract because they want to keep the speed of light constant?

For nonphysicists, these weird, counterintuitive effects of relativity proved an unsurmountable obstacle, because Einstein asks us to believe these effects without giving us any mechanical, intuitive explanation. Even physicists supportive of relativity, such as Arnold Sommerfeld, found this a bit much too much to swallow. After reading Einstein's derivations, Sommerfeld complained to Lorentz: "As ingenious as they are, it seems to me that there is something almost unhealthy in their nonconstructive and unvisualizable dogmatics. An Englishman would hardly have put forth such a theory... I hope you will be able to breathe some life into this ingenious conceptual framework" [21].

The fact is that both time dilation and length contraction have simple physical explanations in terms of the laws of mechanics and electrodynamics (although we really need quantum mechanics for the analysis of the oscillations and sizes of atoms). It is not at all difficult to prove from the laws of relativistic dynamics that clocks, such as atomic clocks, slow down, and solid bodies, such as crystal lattices, contract when moving at high speed [22].

The other surprise in the Lorentz transformations is the relativity of synchronization—but even that has a simple explanation in term of clock transport [23]. Clocks in a moving reference frame are desynchronized relative to our, stationary, reference frame, because slow clock transport in the moving reference frame is not slow in our reference frame, and the extra time dilation accumulated by the transported clock leads to a breakdown of synchronization from the viewpoint of our reference frame. A simple calculation confirms that the transport process accumulates the correct time delay: Suppose that one clock remains at the origin of the moving reference frame and therefore moves at constant speed V, and the other clock is slowly transported with speed $V + \delta V$ until it is separated from the origin by a displacement Δx (measured in our reference frame). If this process takes a time Δt, the two clocks will display a time difference

$$\tau_2 - \tau_1 = \sqrt{1-(V+\delta V)^2/c^2}\Delta t - \sqrt{1-V^2/c^2}\Delta t \simeq \frac{-V(\delta V \Delta t)/c^2}{\sqrt{1-V^2/c^2}} = \frac{-V\Delta x/c^2}{\sqrt{1-V^2/c^2}} \qquad (1)$$

Thus, after the transport is completed, at any given instant of t time, the transported clock displays a time shift $\Delta t' = -(V\Delta x/c^2)/\sqrt{1-V^2/c^2}$ relative to the clock at the origin, which agrees exactly with the time shift calculated from the Lorentz transformation equation for t' .

What these physical explanations of the length contraction and time dilation tell us is that there is nothing magical or teleological about the Lorentz transformations. The Minkowski metric indicates a new geometry for spacetime, but this geometry is not produced by magic, but by the laws of physics. It is not some kind of abstract construct contrived to keep the speed of light constant, as Einstein would have us believe. Poincare once confronted Einstein at one of the Solvay Conferences and demanded to know what mechanics Einstein was using to reach his conclusions about time dilation and length contraction. Einstein dismissed the question; he said "no mechanics" [24]. Poincaré couldn't believe that a physicist would say something like that, and he didn't bother to reply.

After dealing with the time dilation and the length contraction, Einstein's 1905 paper deals with the transformation formulas for electric and magnetic fields, transformation formulas for energy in e.m. waves, aberration, and Doppler shift. The results for electric and magnetic fields had already been obtained by Lorentz, a year earlier; the other results were new.

And then, in the final section of his paper, Einstein formulates new, relativistic equations of motion for a moving charge—called an electron—being accelerated by electric and magnetic

fields. He correctly calculates the so-called longitudinal mass which indicates the effective inertial resistance for acceleration in a direction parallel to the velocity. But he makes a bad mistake in the "transverse mass," for acceleration in a direction perpendicular to the velocity, for which he obtains the mistaken formula

$$Transverse\ mass = \frac{m}{1 - v^2/c^2} \tag{2}$$

Not all of Einstein's mistakes were fruitful, and this was simply a silly mistake, of no redeeming artistic value. In a footnote added years later to a reprint of his paper, Einstein sheepishly admitted that this mistake was not "advantageous." What makes this mistake all the more surprising is that the expressions for transverse and longitudinal mass were well known in the physics literature. Lorentz had used them in his paper on relativistic electrodynamics a year earlier, where he expressed the mass of the electron in terms of its electrostatic self energy and stated the correct velocity dependence for the transverse mass [25],

$$m_2 = \frac{e^2}{6\pi c^2 R} \frac{1}{\sqrt{1 - v^2/c^2}} \tag{3}$$

But Einstein had failed to notice Lorentz's paper.

Einstein's mistake was immediately spotted by Max Planck, who read Einstein's paper soon after it was published (maybe even before it was published, because Planck was editor of *Annalen der Physik*, where had Einstein sent his paper for publication). Planck thought Einstein's paper very impressive, and he sympathized with the top-down approach, perhaps because in his own work on quantization of black-body radiation he had taken the same top-down approach—he had postulated the quantization of energy and deduced the radiation law, just as Einstein had assumed a universal speed of light and deduced the Lorentz transformations and their consequences. Einstein's adoption of the speed of light as a universal constant also pleased Planck, who was very interested in universal constants and their role in physics.

Planck then proceeded to remodel and correct Einstein's treatment of dynamics. With an elegant Lagrangian formulation of relativistic mechanics, Planck recalculated Einstein's transverse mass, and he found and published the correct formula for the transverse mass [26]. Unfortunately, today hardly anybody remembers this contribution of Planck to relativity—it was the first of many papers on relativity by authors other than Einstein to appear after 1905.

3 EQUIVALENCE PRINCIPLE AND REDSHIFT

I now will turn to general relativity and dissect the main mistakes in Einstein's development of general relativity. Like the development of special relativity, Einstein's development of general relativity rests on several great mistakes. These mistakes served him as stepping stones to his final wonderful discovery of curved spacetime.

For Einstein, the key to general relativity was the principle of equivalence of acceleration and gravitation which he discovered in 1907. As he later described it in a lecture at Kyoto University: "I was sitting in my chair at the patent office at Berne. Suddenly I had an idea: when a person is in free fall, he does not feel his own weight. I was amazed. This simple thought experiment made a deep impression on me. It led me to a theory of gravitation" [27]. His sister Maja claimed that what triggered Einstein's sudden idea was the fatal accident of a roofer who slipped from one of the roofs near Einstein's apartment. That story sounds too good to be true, but the roofs of Berne are old and slippery, and unfortunate accidents are plausible.

Einstein saw in this equivalence principle a relativity of acceleration. He thought that physicists placed in an accelerated reference frame cannot be sure whether the effects they feel and observe are due to the pseudoforces associated with the acceleration or to the presence of a constant gravitational field. As he described it in one of his lectures, imagine that two physicists wake up from a drugged sleep and find themselves in a closed box, in which they observe that bodies released in midair fall to the floor with a universal acceleration g. What can the physicists conclude from this? One of them concludes that the box sits on the surface of a planet, whose gravitational attraction produces the acceleration g. The other concludes that the box is nowhere near any planet, and is instead being accelerated by some external propulsion mechanism. And Einstein asks, "Is there any criterion by which the two physicists can decide who is right?" and he answers, "We know of no such criterion" [28].

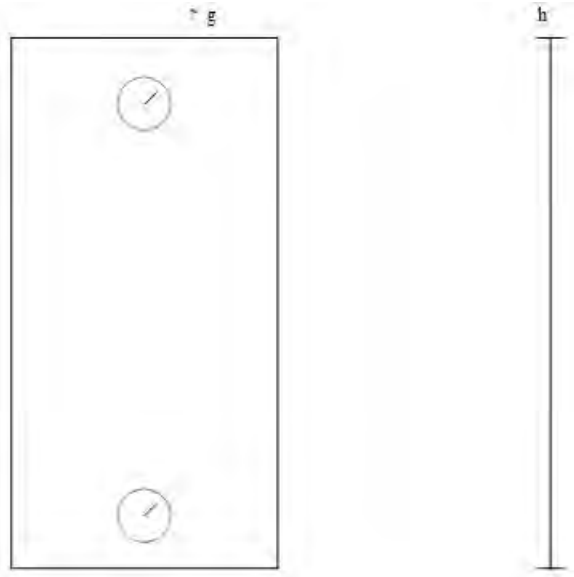
With the wisdom of a hundred years of hindsight, I can give you several such criteria. One obvious violation of the equivalence of acceleration and gravitation is found in the tidal forces generated by gravitational fields, which are absent in a (linearly) accelerated reference frame. These tidal forces arise from gradients in the gravitational field. In Newtonian gravitation these gradients can be zero only under exceptional circumstances; in Einstein's theory of general relativity, they can *never* be zero (they correspond to the components of the Riemann tensor, some of which must be different from zero if the spacetime is curved). It might be argued that such tidal forces become small and insignificant when the region accessible to the experimenter is made extremely small, but even in very a small region, tidal effects remain observable with sensitive equipment. For example, the GOCE satellite of the European Space Agency, in orbit and in free fall, makes high-precision measurements of the tidal gravitational field of the Earth with a differential-accelerometer only a few cm across. In fact, the GOCE satellite measures components of the Riemann tensor (R^k_{00l}) that is, it measures the curvature of spacetime.

Another example is provided by the Stanford Gravity Probe B experiment completed in 2004-2006, which used gyroscopes in a satellite orbiting the Earth to detect the precession caused by a general-relativistic coupling between the spins of the gyroscopes and the Earth. In this experiment, Francis Everitt and his fellow experimenters had to go to extraordinary lengths to avoid interference from tidal forces. They had to manufacture their gyroscopes as perfectly round spheres, to within $\pm 10^{-6} cm$. Even a small deviation from roundness would have permitted the tidal forces to exert torques on the gyroscopes and generate a much larger precession than what the experimenters were looking for, so this experiment would have detected tidal forces rather than spin-spin coupling. Note that the tidal-torque precession is independent of the size of the gyroscopes. We cannot eliminate the tidal-torque precession by using smaller gyroscopes—other things being equal, the precession rate depends on the shape of the gyro, but not on the size.

Besides the obvious troubles with tidal effects, the equivalence principle has a fundamental inability to deal correctly with the propagation of light—the principle works, more or less, for slow-moving particles, but it mishandles fast-moving particles and light, unless we adopt curved spacetime. For instance, we might try to calculate the deflection of light in a gravitational field by beginning with the deflection that occurs in an accelerating elevator. Obviously, in such an elevator, a ray of light moving from one side to the other will deflect downward by some amount relative to the elevator. But, quantitatively, the calculated amount is wrong—the deflection calculated in the elevator is only half as large as the actual result in a gravitational field calculated from Einstein's general relativity (this is the infamous factor of two by which Einstein's first deflection calculations differed from his final result). Thus, the equivalence principle fails by a factor of two, and not because of any tidal effect.

The equivalence principle is capricious and unreliable. The equivalence principle works when it works and doesn't when it doesn't—you have to apply it with caution. Some violations of the principle of equivalence were already mentioned by Arthur Eddington in 1923 in his The

Figure 2: An elevator accelerating upward with two clocks for testing the frequency shift of a lightwave traveling a distance h from floor to ceiling.



Mathematical Theory of Relativity, the first textbook on general relativity. Eddington gave a dismissive opinion about the equivalence principle:

"It is essentially a hypothesis to be tested by experiment as opportunity offers. Moreover, it is to be regarded as a suggestion, rather than a dogma admitting no exceptions. Clearly there must be some phenomena... which discriminate between a flat world and a curved world; otherwise we could have no knowledge of world curvature. For these the Principle of Equivalence breaks down... The Principle of Equivalence offers a suggestion for trial, which may be expected to succeed sometimes, and fail sometimes" [29].

Unaware of these troubles, in 1911 Einstein applied the equivalence principle to a Gedankenexperiment involving a perfectly uniform gravitational field, in which tidal effects are not an issue. [30] He considered a light source located on the surface of the Earth or some other gravitating body, and asked what happens to the frequency of a light wave emitted upward. To find out, he replaced the gravitational field by a reference frame accelerating upward and examined the propagation of light in this reference frame. This Gedankenexperiment is known to all students of relativity, and the accelerating reference frame is usually visualized as an elevator (see Fig. 2). While the light wave travels from the floor to the ceiling, the upward acceleration of the elevator increases the speed of the clock at the ceiling relative to the initial speed of the clock at the floor by about gh/c where g the acceleration of the elevator (equal to the acceleration of gravity the elevator is intended to mimic), h its height, and c the speed of the light wave. Upon arrival at the ceiling, the light wave will then suffer a Doppler shift $\Delta\nu = -(gh/c^2)\nu$. Relying on the equivalence principle, Einstein therefore concluded that a light wave propagating upward in a gravitational field should suffer the same frequency shift. This is the gravitational redshift, also called the gravitational time dilation, because we interpret it as a slowing of clocks in a gravitational field. As a corollary of this gravitational time dilation, Einstein concluded that the speed of light is lower in a gravitational field and that light rays suffer a deflection when passing near a mass.

There is something puzzling about this deduction of the gravitational redshift. The calculation relies on what seems to be no more than simple Newtonian physics, that is, the Doppler-shift formula and the equivalence of gravitational and inertial effects, both of which seem innocuous and trivial. In Newtonian physics, it is immediately evident that a constant gravitational field can be replaced by an accelerated reference frame; the gravitational force in the former is identical to the acceleration pseudoforce in the latter. And yet something is fishy about Einstein's calculation, because, according to Newtonian physics, in a static gravitational field, a propagating light wave cannot acquire a frequency shift, not even if the speed of the wave is somehow altered by the gravitational potential. An alteration of speed would mean that the gravitational field behaves like an optical medium with a position-dependent index of refraction—in such a medium the wavelength of a light wave changes, but the frequency remains constant. This is why we have to interpret the gravitational redshift as a time dilation of the clocks, rather than as an inherent frequency change of the propagating wave. But nowhere in his calculation does Einstein seem to introduce anything about any time dilation, so why does a time dilation of clocks emerge in the end, as if by a trick of magic? Where does Einstein trick us?

The answer to this puzzle is that Einstein actually introduced the special-relativistic time dilation implicitly, by a sleight of hand. He used a mixture of Newtonian and relativistic physics—he took the equivalence principle from Newtonian physics and he took the postulate of a constant, universal speed of light from relativistic physics. He assumed that each wave pulse emitted from the floor of the elevator has the same speed as the preceding pulse, whereas in Newtonian physics we would have to assume that the wave is carried along by the medium in the elevator, and that the wave speed is constant relative to the floor, but increases relative to the inertial reference frame within which the elevator accelerates. The constant speed of light hinges on the special-relativistic time dilation of clocks or, more precisely, it hinges on the relativity of synchronization, which is a consequence of this time dilation (as shown by the clock-transport argument in part I of this lecture).

We can understand the crucial role of relativistic physics in the derivation of the redshift more clearly if we replace the light wave by a sound wave or by an evenly-spaced sequence of BB pellets fired from a BB gun. Each pulse of sound or each BB pellet has the same speed relative to the elevator, but between emission of one pulse and the next, the elevator increases its speed, and therefore, relative to the inertial reference frame within which the elevator accelerates, each pulse of sound or each BB has a slightly higher speed than the preceding one. By taking into account the Newtonian addition law for velocities, we then readily find that the frequency shift is reduced to zero—there is no frequency shift at all [31]. And, by the equivalence principle, we would then conclude that there is no frequency shift for a sequence of sound or pellet signals in a gravitational field.

This makes it clear that to obtain a redshift we need to consider relativistic corrections to the Newtonian calculation (if we need more prompting, the factor $1/c^2$ of in Einstein's redshift formula actually gives us a strong clue that relativistic effects play a role). And, indeed, if instead of the Newtonian addition law for velocity, we use the special-relativistic velocity-addition law, then the Gedankenexperiment with sound pulses or BB pellets yields the expected redshift, exactly as in Einstein's calculation with light waves. For the sake of simplicity, let's assume that the speed u of the sound pulses or BB pellets is much larger than the increment of speed of the elevator during the travel time of the signal. The first pulse is launched at time $t = 0$ and reaches the ceiling at time $t_1 \simeq h/u$ [32]. The next pulse is launched at time $t = \Delta\tau_1$ and reaches the ceiling at time $t_2 = \Delta\tau_1 + h/(u' - g\Delta\tau_1)$ where is the speed of the elevator at launch and is the pulse speed relative to our original reference frame, according to the relativistic combination law for velocities, $u' = (u + g\Delta\tau_1)/(1 + ug\Delta\tau_1/c^2)$ [33]. Ignoring terms of order g^2, we then find

$$t_2 = \Delta\tau_1 + \frac{h}{u} + \frac{gh}{c^2}\Delta\tau_1 \tag{4}$$

The time difference between the arrivals times of the two pulses is therefore

$$\Delta\tau_2 = t_2 - t_1 = (1 + \frac{gh}{c^2})\Delta\tau_1 \tag{5}$$

This redshift formula for the pulse periods agrees with Einstein's redshift formula for the frequencies.

The preceding calculation establishes an important point about the derivation of the gravitational redshift: it arises from an inconsistent mix of Newtonian physics for the equivalence principle and the relativistic addition law for velocity, applied either to light or to sound pulses or pellets, or any other signaling method. The gravitational redshift cannot be derived from purely Newtonian physics [34].

Nor can the gravitational redshift be derived from purely relativistic physics. Although relativistic physics yields the desired expression for the redshift in the accelerating elevator, this leaves a gap in the derivation—in relativistic physics the equivalence principle cannot be deduced from the laws of mechanics, as it can in Newtonian physics. The trouble is that the relativistic generalization of the gravitational force always includes a dependence on the particle velocity, which leads to different accelerations for particles of different velocities. For instance, Lev Okun found that the simplest relativistic generalization of the gravitational force has a strong dependence on the direction of the velocity [35]. For an ultrarelativistic particle moving in a tangential direction, the force has a component in the radial direction and also a component in the tangential direction, whereas for a low-speed particle the force is, of course, purely radial. This discrepancy indicates that the equivalence principle is not valid.

More generally, it is easy to prove that the equivalence principle can never be valid in a relativistic theory of gravitation in flat spacetime. The proof is by contradiction. Suppose the equivalence principle is valid, so all particles have the same gravitational acceleration at any given point, and light propagates with its standard speed relative to these freely falling particles. Then the redshift found in an accelerated reference frame requires a corresponding redshift in a gravitational field, which tells us that clocks in a gravitational field run slow. Furthermore, the equivalence principle tells us that the local speed of light has its standard value $c_0 = 2.99... \times 10^{10}\ cm/s$ when measured by the local, slow, clocks. Following Einstein, we can then conclude that the speed of light must be lower than c_0 when measured by "normal" clocks, that is, the speed of light must decrease in a gravitational field. This leads to a contradiction when we consider an ultrarelativistic particle, of initial speed almost equal to c_0 falling radially downward in the gravitational field of a mass, into regions of stronger and stronger fields. Since this particle must obey the decreasing speed limit set by the speed of light, the particle must decelerate, whereas the equivalence principle demands that the particle must *accelerate*, like a slow-moving particle. This logical contradiction proves that the equivalence principle cannot be valid.

And this raises a troublesome question: Did Einstein fail to notice the logical contradiction between his equivalence principle and the decreased speed of light? Or did he notice, but preferred to keep silent? If Einstein's nondisclosure of the conflict between his picture of a gas of quanta and Planck's law provides a precedent, is would seem that sometimes he preferred to remain silent.

General relativity sidesteps this contradiction in the equivalence principle by exploiting curved spacetime. The distinction between locally measured distances and times in the curved spacetime vs. coordinate distances and times permits us to have a constant local speed of light and

nevertheless a decreasing coordinate speed—we can have our cake and eat it too. Thus, the equivalence principle (except for tidal forces) is consistent with general relativity and curved spacetime, though not with special relativity and flat spacetime [36] .

Can anything be salvaged from the ruins of Einstein's derivation of the gravitational redshift via the equivalence principle? We cannot salvage a quantitative, exact result for the redshift, but we can salvage a qualitative, order-of-magnitude result. Einstein was not entitled to assume that the equivalence principle applies to light or to fast-moving particles—he had no evidence for that, and, as shown above, any attempt to stipulate that the equivalence principle is valid for fast-moving particles leads to a contradiction. He was entitled to assume that it applies approximately to slow-moving particles, but, because of the essential role of relativistic corrections in the derivation of the redshift in an accelerating reference frame, he should have anticipated the possible existence of similar small relativistic corrections in the equivalence principle, which might increase or decrease the redshift. Thus, at most, Einstein could legitimately claim that the gravitational redshift is of the order of magnitude $\Delta\nu = -(gh/c^2)\nu$. The sign of the redshift wouldhave been left undetermined by such an order-of-magnitude estimate, so Einstein could not have been sure whether light slows down or speeds up in a gravitational field.

The absence of a legitimate derivation of the redshift from the equivalence principle weakens the physical motivation that Einstein sought to give his theory of gravitation. But this would still have left him with enough clues to pursue the development of gravitational theory by the same path he followed from 1911 onward, although he would have had to admit that his calculation of the redshift and his subsequent calculation of the deflection of light were only order-of-magnitude estimates. Since his first calculation of the deflection was actually in error by a factor of two, this would not have made much difference to the historical developments that led to the first, failed, German attempt at a measurement of the light deflection in 1912 and, later, to the successful British attempt in 1919.

Of course, the defects in Einstein's derivation of the redshift do not mean that his formula for the gravitational redshift is wrong—these defects merely mean that the formula cannot be derived by the seductively simple argument proposed by Einstein in 1911, an argument uncritically imitated in just about every textbook on relativistic gravitation, introductory or advanced [38].

The inconsistent mix of Newtonian and relativistic physics that Einstein used in his derivations of the gravitational redshift was one of his great mistakes. Einstein was as obsessive and dogmatic about the equivalence principle as he was about his stipulation for the one-way speed of light, and he failed to see that the equivalence principle is self-evident only when it involves no more than the equality of rates of free fall of small test masses with low speeds, that is, when it reduces merely to the equality of gravitational and inertial masses.

Years ago, the Irish physicist J. L. Synge said "The Principle of Equivalence performed the essential office of midwife at the birth of general relativity," and he added, " I suggest that the midwife be now buried with appropriate honours and the facts of absolute space-time be faced" [39], by which Synge really meant the facts of absolute curved spacetime. I think Synge was right, and it would be best to forget about the equivalence principle, except in a historical context. Late in his life Einstein said about Mach's principle "Actually, one should no longer speak of Mach's principle at all" [40]. We can say the same about the equivalence principle.

4 GENERAL RELATIVITY

As used by Einstein in the derivation of the redshift, the equivalence principle was a mistake, but it was his greatest and most wonderful mistake, and it led him to his greatest, most wonderful discovery—curved spacetime. We now understand that the time dilation actually implies that

spacetime is curved—the time part of space-time is "shorter" in a gravitational field than outside of the field. We can have a redshift in a static gravitational field only if we have a curved spacetime. But when Einstein published his paper on the equivalence principle in 1911, he did not yet understand that. He was aware that his redshift result means that clocks at a deeper gravitational potential run slower, but he vaguely speculated that that this arose, somehow, from a reduced speed of light—he was not yet ready to imagine a curved spacetime. As Emilio Segré liked to say, "When the mind is not prepared, the eye does not recognize" [41].

Einstein finally reached the conclusion that spacetime must be curved in 1912, by a different path. In the spring of that year, while still at the University of Prague, he published a paper [42] exploring his idea that the speed of light in a gravitational field is slower by an amount depending on the gravitational potential, $c = (1 + \Phi/c_0^2) \times c_0$. He had deduced this expression from the gravitational time dilation, by assuming that the speed has the standard value $c_0 = 2.99...\times 10^{10}\ cm/s$ when measured by the local, slow, clocks and is therefore lower when measured by "normal" clocks (his expression for the speed is actually wrong; the actual dependence of the speed of light on the potential is twice as strong, that is, a factor of 2 is missing). At the end of this paper, in a short appendix added at the last moment in the proofs, he draws attention to a curious fact about the equation of motion of a particle that he had derived from his relationship between the speed of light and the gravitational potential: the equation of motion can be expressed by a principle of extremum action, involving the variable speed of light,

$$\delta \int \sqrt{(1 + \Phi/c_0^2)c_0^2 dt^2 - dx^2 - dy^2 - dz^2} = 0 \tag{6}$$

This equation coincides precisely with the equation for a geodesic in a curved spacetime, with a metric tensor $g_{00} = (1 + \Phi/c_0^2)$ [43]. The last sentence of the appendix Einstein says that he suspects that his equation has a much deeper meaning and that it reveals how the equations of motion are to be constructed in general.

Here Einstein fell just short of recognizing that his equation is a geodesic equation, but this revelation came to him soon thereafter. He remembered that he had seen such equations for the extremum length of geodesics in some lectures on curved spaces he had attended at the Polytechnic. This gave him the missing link, and Einstein suddenly saw the real meaning of gravitation: there are no gravitational forces; there is only a curved spacetime in which particles move on the straightest possible wordlines. As he later reported in his Kyoto lecture: "...I suddenly realized that Gauss's theory of [curved] surfaces holds the key for unlocking this mystery. I suddenly remembered that Gauss's theory was contained in the geometry course given by Geiser when I was a student. I realized the foundations of geometry have physical significance" [44].

And from that time on he relentlessly pursued the goal of a geometrical theory of gravitation. Years later, after his success with general relativity brought him worldwide fame, one of his Zurich friends told him he had found the certainty of Einstein's convictions about gravitation almost frightening: "Your confidence , the confidence of your thinking, ..., at the time you when were with us, is for me a tremendous psychological experience. You were so certain, that your certainty had for me an overwhelming effect" [45]. This absolute certainty in Einstein's convictions is also revealed by the famous comment he made to a student when he received the first reports of the observational confirmation of his prediction for the deflection of light. He said: "I always knew the theory was right." The student asked him, What if the measurements had contradicted your theory? And Einstein gave the grand reply: "Then I would have felt sorry for the Dear Lord. My theory is right anyway" [46].

Einstein's path from his introduction of curved spacetime in 1912 to his final theory of general relativity in 1915 was a lengthy and arduous effort. Einstein had little knowledge of the

mathematics of curved spaces; there were few books from which he could learn this material, and those that were available were written in the usual impenetrable jargon of mathematicians, and with awkward notation. He begged for help from his friend Marcel Grossmann: "Grossmann you must help me or I'll go crazy" [47]. But even Grossmann was not sufficiently familiar with Riemannian geometry and did not understand the critical relationship between the differential identities of the Riemann tensor and the conservation laws in the equations that he and Einstein tried to construct. This led to a series of proposals for theories that all had to be quickly abandoned. As Einstein himself admitted later: "The series of my publications on gravity is a chain of wrong turns" [48]. Planck advised Einstein to give up the attempt, "As an older friend," he said, "I must advise against it... In the first place you won't succeed; and even if you succeed, nobody will believe you" [49].

In the end, in 1915, Einstein succeeded brilliantly. And in the next year, he published a long and careful exposition of his theory in Annalen der Physik in 1916. The paper consists of an introduction that tries to lay the physical foundation for the theory, followed by a careful exposition of the essential aspects of Riemannian geometry, then the new field equations for the gravitational field, and finally applications to light deflection and the perihelion precession of Mercury.

Oddly, the discussion of the physical foundations of the theory does not mention the 1912 argument that initially led him to the idea of a curved spacetime. And, equally oddly, this argument has been rarely used in later textbooks on general relativity. The exception is an early book of by Tullio Levi-Civita (1923), an Italian mathematician and expert on Riemannian geometry, who gave a simple argument based on that of Einstein, but leading directly from the Newtonian potential to curved spacetime [50]. Levi-Civita simply wrote the extremum principle for a particle moving in a Newtonian gravitational field as

$$\delta \int c^2(1 - v^2/2c^2 + \Phi/c^2)dt = 0 \tag{7}$$

where the integrand differs from the usual Lagrangian for motion in a gravitational field only by an irrelevant minus sign and factor of c^2. Taking advantage of the small magnitude of velocity and Newtonian potential compared with c, he rewrote this approximately as

$$\delta \int c^2 \sqrt{(1 - v^2/2c^2 + 2\Phi/c^2)}dt = \delta \int c^2 \sqrt{(1 - 2\Phi/c^2)dt^2 - dl^2/c^2} = 0 \tag{8}$$

which is essentially the same as Einstein's extremum principle, and corresponds to motion in a curved spacetime geometry, with a metric tensor whose 00 component depends on the Newtonian potential, and therefore depends on space (this is always a curved spacetime geometry, except when $g_{00}\ \alpha\ x^2$).

In his 1916 paper, Einstein does not use the straightforward and conclusive argument of 1912. Instead, he tries to use the equivalence principle. He considers a rotating turntable, like a Merry-go-round, and he claims that the geometry of this turntable is a curved space, because meter sticks laid along the circumference contract by the usual length contraction of special relativity, whereas meter sticks laid along the radius do not. This makes the ratio of the measured circumference larger than the measured radius, that is, circumference $> 2\pi r$. But this is simply a misconception: the meter sticks at different locations around the circumference are instantaneously in different inertial reference frames, and it is evidently incorrect to add lengths measured in such different reference frames. The apparent curved geometry is an artifact resulting from a bad choice of measurement procedure. If we use meter sticks at rest, floating just above the circumference of the turntable, we will of course find that the geometry is flat. The transformation from coordinates at rest to coordinates in rotation is merely a coordinate

transformation, which cannot change the geometry, that is, the curvature tensor is zero, no matter what coordinate system we use to calculate it.

But Einstein, and quite a few of his followers, accepted this Merry-go-round argument and concluded that the centripetal acceleration of the turntable produces a curved geometry, and if so, by the equivalence principle, the gravitational field of a mass should also produce a curved geometry. Einstein would have done better to motivate the curved spacetime geometry by the extremum principle that first gave him the key to curved spacetime in 1912. Or he could have used the gravitational time dilation that he had extracted from his elevator Gedankenexperiment.

Einstein called his theory "general relativity" for two reasons: he thought that the behavior of physical systems in an accelerated reference frame is indistinguishable from their behavior in an unaccelerated reference frame placed in a gravitational field; and he thought that by writing the equations of physics in a form that was valid in all conceivable coordinate systems he was giving them some kind of new relativity, more general than the relativity of the special theory. He called this the covariance principle of the equations of physics. The first of these reasons was an outright mistake—acceleration and gravity can be distinguished by suitable experiments, as I have already pointed out. And the second of these reasons reflects a misconception about the meaning of relativity. Validity of all laws of physics in all coordinate systems is not a principle of relativity—it is a triviality. It's like saying that an elephant remains an elephant when you express its height in meters instead of centimeters. The mathematician Erich Kretschmann soon pointed out to Einstein that, of course, all laws of physics can be expressed in all conceivable coordinates, even Newton's laws can be expressed in all coordinates, although they then look very messy (for instance, in rotating coordinates, they acquire centrifugal and Coriolis terms).

The real meaning of Einstein's covariance principle was hidden in a tacit assumption that Einstein failed to state, but took for granted: not only can the laws of physics be expressed in all conceivable coordinates, but when you express the laws in the special coordinates that correspond to a local geodesic reference frame —that is, a freely falling reference frame [51] —at and near one point, then the laws reduce to those of special relativity. The real content of Einstein's covariance principle lies in this added condition. Obviously, the added condition imposes severe restrictions on how matter can and cannot couple to the gravitational field; in fact, it completely determines the details of these couplings.

The added condition makes Einstein's covariance principle into a principle of gauge invariance, analogous to the gauge invariance of electrodynamics. I don't want to burden you with the mathematics of gauge invariance, but here is a simple physical explanation. Suppose we place a small Faraday cage at some point in an electric field. Then in the interior of the cage, the electric field will be reduced to zero, but the electric potential associated with the external electric field will remain different from zero. However, any experiment we perform inside the cage will be totally uncoupled from the external electric field, and will be totally unaffected by the potential. The potential in the cage is different from zero, but it is merely a physically irrelevant additive constant—this is the principle of gauge invariance for the potential. (It has among its consequences the conservation of electric charge, as shown by a neat, elementary, argument of Wigner's [52].

In a completely analogous manner, suppose we consider a small laboratory in free fall in a gravitational field. The laboratory plays the role of a Faraday cage for gravitation. Within the laboratory, the gravitational field disappears (more or less), and there only remains the gravitational potential—or, more precisely, the metric tensor that plays the role of potential in Einstein's theory. Einstein's covariance principle, with the additional tacit assumption included, tells us that the laws of physics in the freely falling laboratory are the same as those of special relativity, and therefore any experiment we perform in the laboratory will be uncoupled from the external gravitational field, and will be totally unaffected by the gravitational potential. Thus,

the covariance principle is a principle of gauge invariance. In essence, this invariance principle can be regarded as a general mathematical implementation of the equivalence principle, for all particles (whether fast-moving or slow-moving), for light, and for all the laws of physics [53].

The modern view of covariance is somewhat broader and permits theories of gravitation that go beyond Einstein's. Instead of insisting that the laws of physics in a freely falling reference frame are exactly those of special relativity, we are today willing to accept that they differ by terms involving the Riemann tensor, that is, we regard nonminimal couplings as consistent with covariance. Some of the theories to be presented at our conference rely on this broader interpretation of covariance. Vladimir Fock, a Soviet expert on relativity, severely criticized Einstein for calling his theory general relativity; he claimed that in general relativity there can be no relativity, and said "However paradoxical it may seem, Einstein himself... showed such a lack of understanding when he named his theory" [54]. John Wheeler proposed to replace the name of the theory by "geometrodynamics," in analogy with electrodynamics. But the name general relativity has stuck, and nothing can now be done about that. It is probably best to ignore this unfortunate choice of name and remember that many parents give totally silly names to their offspring—and custom grants them the right to do so.

The gauge invariance arising from the principle of covariance has broad physical consequences: it implies conservation laws for energy momentum and other "conservation" laws, and in Einstein's theory it implies the exact equality of inertial mass and gravitational mass, even for systems that contain substantial amounts of gravitational self-energy [55] . The gauge invariance of Einstein's theory also compels gravitational waves to have spin 2, and *only* spin 2. I regard this as the most fundamental physical consequence of covariance. And the argument can be turned around: if we assume that the carrier of gravitational interactions is a spin-2 field, without any spin 1 or spin 0 components, then it must be a tensor field with gauge invariance, and from this we can conclude that the field equations must have covariance and must coincide with the Einstein equations. If Einstein had not discovered his theory in 1915, it would have been discovered sometime around 1930, when gauge transformations and their implications for the spin content of fields came to be understood.

It is ironical that in his mathematical calculations with his field equations in the 1916 paper, Einstein proceeded without general covariance. Although he repeatedly affirmed the general covariance principle, he found it inconvenient to adhere to this principle, and he wrote his equations in a form that is not generally covariant. Today we write these equations in the form $R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = -8\pi G T_{\mu\nu}$, and we call them the "Einstein's equations," but Einstein didn't write them that way. Instead, he adopted the condition $\sqrt{-g} = 1$ and wrote his equations in a more convenient, simplified form $\partial\Gamma^{\alpha}_{\mu\nu}/\partial x_{\alpha} + \Gamma^{\alpha}_{\mu\beta}\Gamma^{\beta}_{\nu\alpha} = -\kappa(T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T)$. But this form of the field equations is not generally covariant, because it is not generally true that $\sqrt{-g} = 1$. It's a case of Do as I say but don't do as I do.

Actually, the equations $R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = -8\pi G T_{\mu\nu}$ were first obtained by David Hilbert, the well-known Göttingen mathematician. And what is more, he announced these equations in a lecture at Göttingen a few days before Einstein. But he then made the mistake of setting $R = 0$, because he was interested only in the gravitational fields associated with electric and magnetic fields. By this mistake, Hilbert left the resurrection of this term to Einstein, a few days later. John Auping has aptly called these alternating forward and backward steps in the approach to the field equations the pas de deux of Einstein and Hilbert [56]. This erratic historical development brings to mind the words that Kepler used about his own road to discovery, "the roads that lead man to knowledge are as wondrous as that knowledge itself" [57].

From a broader perspective, we can see that Einstein's 1916 paper on general relativity suffers from much the same problems as his 1905 paper on special relativity. The physical foundations are shaky and riddled with mistakes, but by his amazing intuition, Einstein arrives at correct,

or almost correct, final results despite of that. He somehow managed to find his way through the fog of his own confusion and reach his goal despite his mistakes.

The crucial mistakes in both papers were based on sudden, inspirational ideas, to which Einstein took an obsessive liking, and which he elevated into dogmas. These were big mistakes, but they also were wonderful mistakes that led Einstein to astounding discoveries. We can say of Einstein what Arthur Koestler said of Kepler: "The measure of Kepler's genius is the intensity of his contradictions, and the use he made of them" [56]. To which I will add, To err is human, and to err greatly is divine... at least sometimes.

References

[1] H. C. Ohanian, Einstein's Mistakes (W. W. Norton and Co, New York, 2008). For brevity, this book will be designated by E.M. hereafter.

[2] E.M., p. 337.

[3] G. Kane, "String theory and the real world," Physics Today, November 2010, p. 39.

[4] E.M., p. 166.

[5] L. S. Feuer, Einstein and the generations of science (Transaction Publishers, New Brunswick, NJ, 1982), p. xiii.

[6] A. Einstein, Annalen der Physik 17, 891 (1905).

[7] S. T. S. Lecky, Wrinkles in Practical Navigation (Philip and Son, London, 1942), p. 44.

[8] For instance, A. Ganot's Éléments de Physique, published in more than a dozen editions, and also available in a German translation. In this book, Roemer's method is discussed in Section 505 (fourteenth edition).

[9] P. Galison, Einstein's Clocks, Poincaré's maps (W. W. Norton and Co, New York, 2003), p. 291.

[10] A. Pais, 'Subtle is the Lord...' (Clarendon Press, Oxford, 1994), p. 173.

[11] J. Auping Birch, Una revisión de las teorías sobre el Origen y la Evolución del Universo (Universidad Iberoamericana, 2009), p. 141.

[12] A. Einstein, Berliner Tageblatt, December 25, 1919 (my translation).

[13] A. A. Martínez, Kinematics: the lost origins of Einstein's relativity (Johns Hopkins Press, Baltimore, 2009).

[14] A symmetry argument of this kind was stated by Wolfgang Pauli in his book Theory of Relativity (Pergamon Press, London, 1958), originally published as an article in the Mathematical Encyclopedia in 1921, and maybe it was known earlier.

[15] E.M., p. 17.

[16] E.M., p. 332.

[17] E.M., p. 2.

[18] H. Minkowski, "Space and Time," Address delivered at the 80th Assembly of German Natural Scientists and Physicians, Cologne, September 1908.

[19] C. Fuentes, La muerte de Artemio Cruz (Penguin Books, New York, 1996), p. 255.

[20] E.M., p. 162.

[21] E.M., p. 97.

[22] W. F. G. Swann, Rev. Mod. Phys. 13, 197 (1941) and J. S. Bell, Speakable and Unspeakable in Quantum Mechanics (Cambridge University Press, Cambridge, 1987).

[23] D. Bohm, The Special Theory of Relativity (Routledge, London, 1996), p. 33. Bohm's argument is stated in terms of clock frequencies, which is an unnecessary complication.

[24] E.M., p. 281.

[25] H. A. Lorentz, Proceedings of the Academy of Sciences of Amsterdam 6 (1904).

[26] M. Planck, Verh. d. Deutsch. Phys. Ges. 8, 136 (1906).

[27] E.M., p. 183.

[28] E.M., p. 185.

[29] Arthur Eddington, The Mathematical Theory of Relativity (Cambridge University Press, Cambridge, 1963), pp. 40, 41.

[30] Annalen der Physik, 35, 818 (1911).

[31] This absence of redshift is immediately obvious if we consider that, in the instantaneous reference frame of the elevator at launch, each successive sound or pellet signal has exactly the same initial conditions, and therefore necessarily has the same travel time to the ceiling. The absence of redshift in Newtonian physics was first noticed by P. Florides and published in the announcement of a lecture [International Journal of Modern Physics A, 20, 2759 (2002)]. However, the announcement gives no details and does not say what conclusions Florides drew from his discovery. Florides has informed me that he expects to publish more about this topic shortly, but has not yet done so.

[32] This assumes that the speed u is large ($u^2 >> gh$) so the pellets are not significantly decelerated during their upward motion and, furthermore, the displacement $gt^2/2$ of the ceiling during the travel time can be ignored.

[33] Note that the crucial term in this expression for u' is the denominator, which represents the deviation from Newtonian physics. The term $g\Delta\tau_1$ in the numerator of u' and the same term in the denominator of t_2 can usually be neglected; in any case, these terms approximately cancel. In Newtonian physics, which corresponds to c^2 tending to infinity, these terms cancel exactly, and the redshift then disappears.

[34] There is another simple derivation of the gravitational time dilation that focuses directly on the role of relativity and dispenses with the imbroglio of signal transmissions. To compare the rate of a clock 2 with the rate of a clock 1, simply drop an auxiliary clock from 2 to 1. Since this auxiliary clock is in free fall, the equivalence principle tells us that it is unaffected by the gravitational field, so we can presume its rate is constant and we can use this clock as a standard of frequency. When the auxiliary clock arrives at clock 1, it will have a speed

$v = \sqrt{2gh}$. Hence clock 1 will have a time dilation factor relative to the auxiliary clock and also relative to clock 2. This is the usual gravitational time dilation factor $\sqrt{1 - v^2/c^2} \simeq 1 - gh/c^2$ for the lower clock relative to the upper. The argument can be easily extended to nonuniform gravitational fields; see H. C. Ohanian and R. Ruffini, Gravitation and Spacetime (W. W. Norton and Co, New York, 1994), pp. 167, 168.

[35] L. B. Okun, "The Concept of Mass," Physics Today, January 1989, p. 31.

[36] A simple calculation shows that in general relativity, in the Schwarzschild geometry, the coordinate speed of a freely falling particle decelerates monotonically if the initial, asymptotic, speed is larger than $c/\sqrt{3}$. A particle of a lower speed first accelerates, and then decelerates when it approaches the Schwarzschild radius (H. C. Ohanian, "Reversed gravitational acceleration of high-speed particles," to be published).

[37] And, I have to confess, also in my own textbook, H. C. Ohanian and R. Ruffini, op. cit. This mistake will be corrected in the next edition, currently in progress.

[38] For instance, T. Padmanabhan, Gravitation, Foundations and Frontiers (Cambridge University Press, Cambridge, 2010), pp. 125-127, gives a very thorough, exact treatment (by Rindler coordinates) showing how the redshift is generated by acceleration of a reference frame. But he then misuses the equivalence principle to obtain the redshift in a gravitational field. In this he commits the same mistake as Einstein: he proves that the equivalence principle is valid for slow-moving particles, and then takes it for granted that the same is true for fast-moving particles or light.

[39] J. L. Synge, Relativity, the General Theory (North Holland, Amsterdam, 1971), p. IX.

[40] E.M., p. 249.

[41] E.M., p. 189.

[42] Annalen der Physik, 38, 443 (1912).

[43] If $\Phi = gx$ then this is the Rindler metric, and not curved.

[44] A. Pais, op. cit., pp. 211, 212.

[45] E.M., p. 256.

[46] E.M., p. 255.

[47] E.M., p. 192.

[48] E.M., p. 198.

[49] E.M., p. 205.

[50] T. Levi-Civita, The Absolute Differential Calculus (Dover, New York, 1977), p. 293; see also W. Rindler, Essential Relativity (Springer, New York, 1977), p. 123.

[51] Geodesic here means $\Gamma^{\mu}_{\alpha\beta} = 0$, that is, geodesic in regard to the Einstein tensor field.

[52] E. Wigner, Symmetries and Reflections (Indiana University Press, Bloomington, 1967), pp. 11, 12.

[53] Invariance can also be stated as a requirement that the laws have the same "form" in all coordinates, so they are general tensor equations and the tensor character is achieved by means of the gravitational field tensor, and no other tensor (that is, there is no other auxiliary prior geometry, or absolute object).

[54] V. Fock, The Theory of Space, Time, and Gravitation (McMillan Co, New York, 1964), p. 8.

[55] H. C. Ohanian, arXiv:1010.5557[gr-qc], October 2010. There is an intriguing exception to the equality of inertial and gravitationsl mass: the inertial mass of a wormhole is twice its gravitational mass, something that would be nice to test if we ever find a wormhole.

[56] J. Auping Birch, op. cit., p. 162n.

[57] E.M., p. 331.

[58] E.M., p. ix.

Putting the standard ΛCDM model and the relativistic models in historical context

John A. Auping

Department of Physics and Mathematics, Universidad Iberoamericana, Mexico City, Mexico.
E-mail: jauping@iwm.com.mx

Abstract. The historical origin of the idea of the apparently missing matter, later labeled as dark matter, is shown to be the use of Newtonian mathematics in trying to explain the rotation velocity of stars in spiral galaxies and galaxies in galaxy clusters. Much attention is paid to Thieu and Cooperstock, who, in using relativistic dynamics, have shown that dark matter can be disposed of as a myth. Dark energy is shown to be an artifact originally hypothesized to explain the apparent recent acceleration of the expansion velocity of the Universe. This apparent acceleration is shown to be an optical illusion that can be disposed of as a myth, once Einstein′s gravitational theory is taken seriously. Much attention is paid to Buchert′s and Wiltshire′s work on explaining the apparent acceleration of the expansion velocity of the Universe without dark energy.
Keywords: dark matter, dark energy, rotation velocity of galaxies and galaxy clusters, expansion velocity of the Universe, Newtonian gravitational dynamics, relativistic gravitational dynamics.

Contents

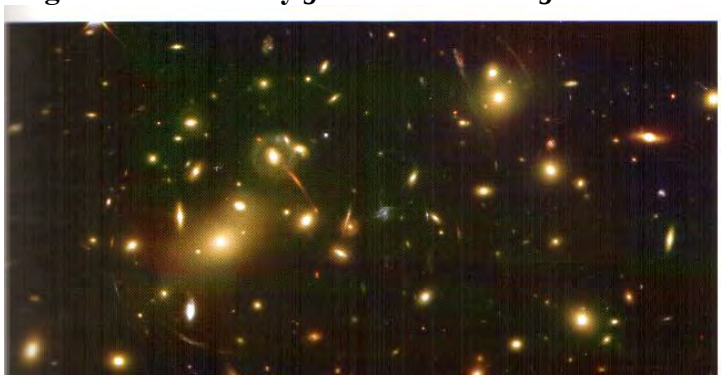
Part 1.- Origin of the speculation on non-baryonic dark matter

1.1.- Dark matter in spiral galaxies

The amount of non-baryonic dark matter is established through the discrepancy between the observed visible mass –which is baryonic— and the total mass calculated from certain effects generated by gravitational fields. The first person alerting to the supposed missing mass in galaxies and galaxy clusters, in 1933, was Fritz Zwicky (1898-1974), a Swiss astronomer working in Pasadena, California. By comparing the redshift of individual galaxies that belong to a cluster with the redshift of the entire cluster, he was able to establish the proper velocity of a galaxy. Thus he could prove that the orbital velocity of galaxies in a cluster is higher than expected if one would only take into account the mass of visible matter (stars and ionized gas) from the point of view of Newtonian gravitational dynamics.[1] Along this same line of reasoning, Rubin y Coyne speculated that the method of establishing the peculiar velocities of galaxies within a cluster through their redshift, serves to reveal "*the relative distribution of dark and luminous matter*" and "*indicates the existence of large amounts of (dark) matter.*"[2]

Zwicky also established that one can determine the total mass of a galaxy through the observation of the curvature of light coming from a star or galaxy that is located behind the Sun or a galaxy cluster. This method, derived from the theory of general relativity, served originally to corroborate this theory. Now that it has been corroborated, one proceeds in the opposite order and the curvature of light allows us to calculate the total mass of a galaxy that is located between a luminous object and the Earth.[3] According to Zwicky, the observation of these effects of gravitational lensing provides us with the most simple and most exact determination of the masses of galaxies.[4]

The arcs of galaxies curved by *gravitational lensing*: the Abell 2218 cluster



[1] Fritz Zwicky, "Die Rotverschiebung von extragalaktischen Nebeln", *Helvetica Physica Acta,* vol. 6 (1933): 110-127; "On the Masses of Nebulae and of Clusters of Nebulae", in: *The Astrophysical Journal,* v. 86 (1937): 217 -46

[2] Vera Rubin & George Coyne, eds. *Large Scale Motions in the Universe* (1988): 262, 101-102

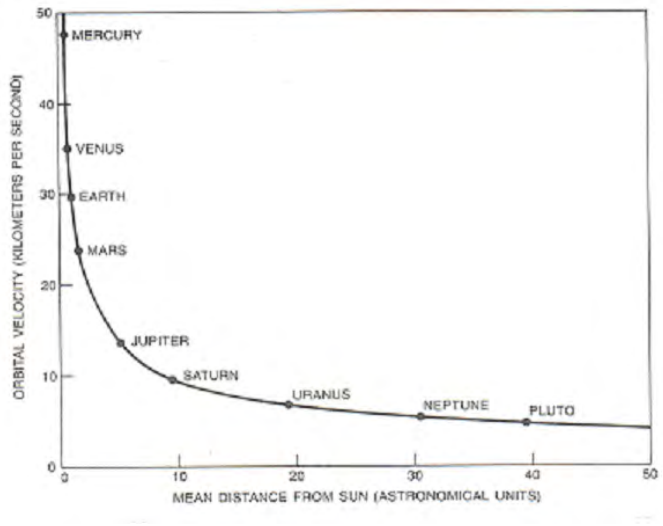
[3] The curvature is $\varphi = 0.5\,g\,l/c^2$ *radians* where l is the distance travelled by light through a gravitational field and g is the gravitational acceleration. The factor g depends directly on the mass of the object that causes the bending of the light. See George Gamov, *En el país de las maravillas. Relatividad y cuantos* (1958): 96

[4] Fritz Zwicky, "On the Masses of Nebulae and of Clusters of Nebulae", in: *The Astrophysical Journal,* vol. 86 (1937): 238. Zwicky denominates "nebulae" what we know today to be galaxies.

Observations have been made of the orbital velocities of stars in spiral galaxies that seem to reveal the existence of a halo of non-baryonic dark matter which extends further than the visible disk of the galaxy, on the condition that we interpret these velocities from the point of view of Newtonian gravitational laws. These laws predict that the acceleration diminishes with the inverse square of the distance of the central mass and that the orbital velocity diminishes with the inverse root of the distance. In a series of 10 important publications, in *The Astrophysical Journal* from 1977 to 1985, Vera Rubin and her team observed about 60 spiral galaxies (20 of type Sa, 20 of type Sb and 20 of type Sc ([note 5]) and reported that the orbital velocity was almost constant, independent of the distance of the center of the galaxy.[6] It is important to distinguish between the astrophysical *observations* of Vera Rubin and her team and the *interpretations* they made of these observations. Two cosmographic observations are beyond any doubt:

1.- *First observation*.- In the solar system, the orbital velocity of the planets diminishes with the inverse root of the distance ($v \propto 1/r$), so the velocity diminishes as the distance increases:

Graph.- First observation: orbital velocity and distance from the Sun in the solar system[7]



2.- *Second observation*: the rotational velocity of stars of a spiral galaxy, increases rapidly at short distance from the galaxy center, then stops diminishing with distance, and remains more or less constant (the curve flattens out). However, the visible mass diminishes rapidly as one moves away from the galaxy center.

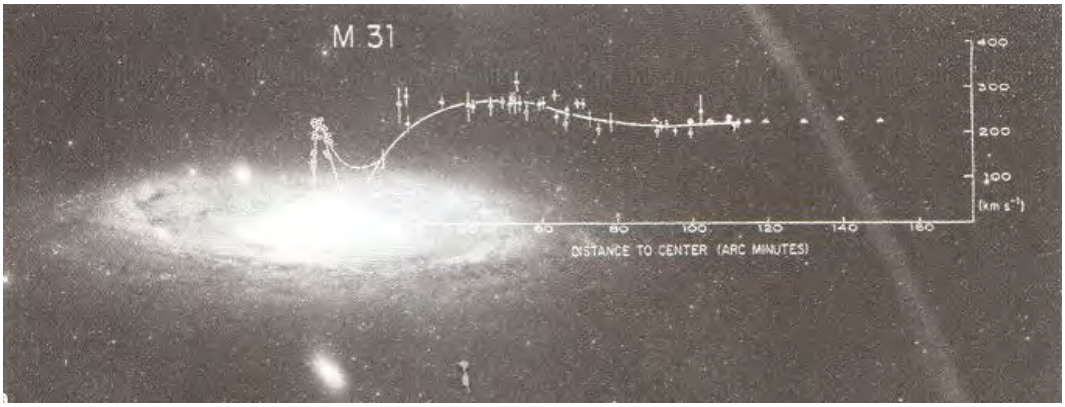
---

[5] Spiral galaxies type Sa have a big center, with the arms close to each other; galaxies type Sb, a smaller center with distinguishable arms; and the type Sc galaxies, an ever smaller center with arms quite separate from each other.

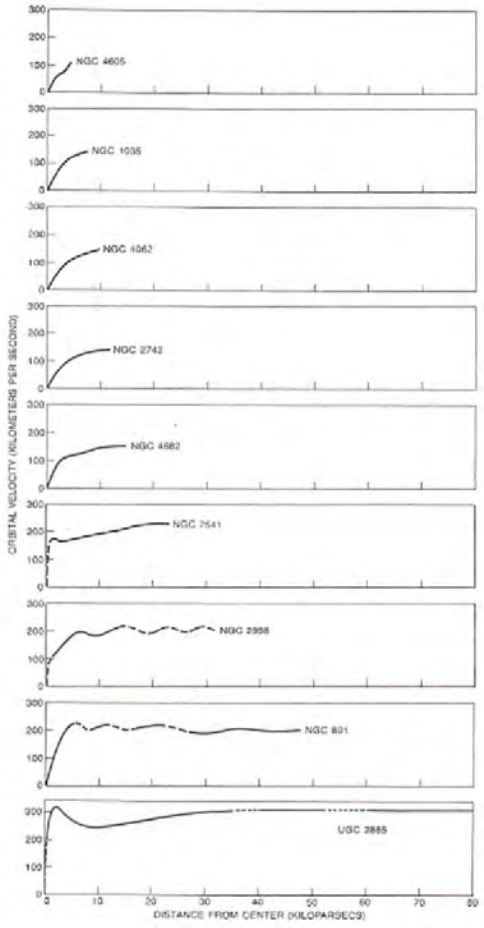
[6] Vera Rubin & Kent Ford *et al.* "Extended rotation curves of high-luminosity spiral galaxies. I The angle between the rotation axis of the nucleus and the outer disk of NGC 3672," *The Astrophysical Journal*, vol. 217 (1977): L1-L4; "II The anemic Sa galaxy NGC 4378," *ibidem,* vol. 224 (1978): 782-795; "III. The spiral galaxy NGC 7217," *ibidem,* vol. 226 (1978): 770-776; "IV. Systematic dynamical properties," *ibidem,* vol. 225 (1978): L107-L111; "V. NGC 1961, The most massive spiral known," *ibidem,* vol. 225 (1979): 35-39; "Rotational properties of 21 Sc galaxies with a large range of luminosities and radii, from NGC 4605 (R=4 lpc) to UGC 2885 (R=122 kpc)" *ibidem,* vol. 238 (1980): 471-487; "Rotation and mass of the inner 5 kiloparsecs of the SO galaxy NGC 3115," *ibidem,* vol. 239 (1980): 50-53; "Rotational properties of 23 Sb galaxies," *ibidem,* vol. 261 (1982): 439-456; "Rotation velocities of 16 Sa galaxies and a comparison of Sa, Sb, and Sc rotation properties," *ibidem,* vol. 289 (1985): 81-104

[7] Original drawing by Vera Rubin in *Scientific American* (1983): 90

Image.- Photo of the spiral galaxy M31 in which Rubin drew the flat rotation curve [8]



Graph.- Second observation: orbital velocity and distance of the center in 9 type Sc galaxies [9]



These two *observations* of real facts in spiral galaxies are *interpreted* by Vera Rubin and her team from the point of view of a cosmological model with Newtonian gravitational dynamics:

[8] Malcolm Longair, *Galaxy Formation* (2008): 67
[9] Original graphs by Vera Rubin in *Scientific American* (1983): 93

1.- *First part of the interpretation*. Rubin and her team start from the assumption that in spiral galaxies the gravitational dynamics operating are Newtonian. According to Newton the gravitational acceleration diminishes with the inverse square of the distance, and orbital velocity diminishes with the inverse root of the distance, as is explained in the next mathematical box.

MATHEMATICAL BOX 1, ORBITAL VELOCITY IN THE CONTEXT OF NEWTONIAN GRAVITATIONAL DYNAMICS

The mathematical reasoning proper of a Newtonian model is the following. According to Newton's second law of movement, the acceleration a is:

$$F = ma = GMm/r^2 \Rightarrow a = GM/r^2 \quad (1)$$

and the acceleration of a body in orbit around a big central mass is:

$$a = v^2/r \quad (2)$$

From (1) and (2) we deduce:

$$GM/r^2 = v^2/r \quad (3)$$

And from (3) we deduce the orbital or rotational velocity:

$$v = \frac{\sqrt{GM}}{\sqrt{r}} \quad (4)$$

which means that the orbital velocity is proportional to the root of the mass and inversely proportional to the root of the distance:

$$v \propto \frac{\sqrt{M}}{\sqrt{r}} \quad (5)$$

Since it is reasonable to assume that the mass of the galaxy is concentrated at its center and diminishes if one moves away from the center, one would expect, according to equation (5) that the rotational velocity v would rapidly decrease if one moves away from the center, since in that assumption with the total mass M (within the range r) gradually stopping to inrease and r increasing linearly, $\sqrt{M}/\sqrt{r}$ would diminish rapidly. The surprise is that one observes the contrary: the orbital velocity, at a certain distance from the center and beyond, remains constant even though we move away from the center. The only way to explain this strange phenomenon is the assumption that the mass, instead of gradually stopping to increase, actually increases linearly with radius, up to a certain, far away distance. For example, at twice the distance from the center, we would have twice the mass. That would explain why the rotational velocity remains constant with distance:

$$if\ M \propto r \Rightarrow \Delta v = 0 \quad (6)$$

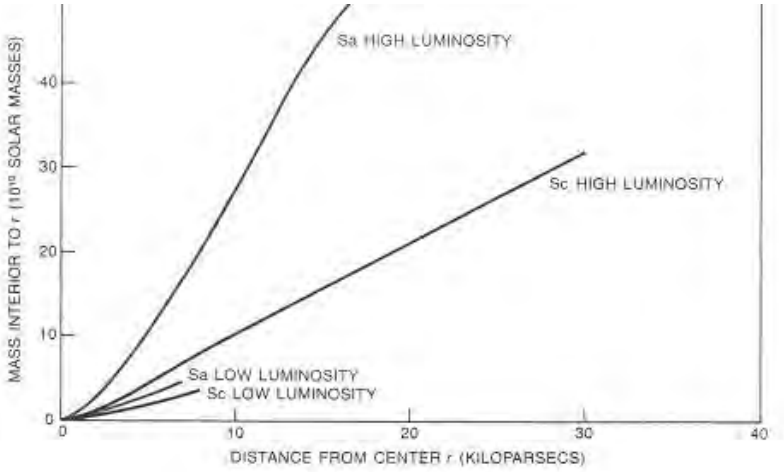
So, within the context of Newtonian gravitational dynamics there is no other option left, but to assume the existence of a supposed halo of non-baryonic dark matter.

2.- *Second part of the interpretation.* The second part of the interpretation of Rubin and her team is the following: *given Newtonian gravitational dynamics*, an apparent incompatibility arises

between the visible galaxy mass (stars and gas), observed through the radiation in some frequency of the electromagnetic specter, and the observed constancy of the orbital velocity. This is interpreted as requiring enormous quantities of additional and invisible mass, that increase linearly with distance from the center, as can be seen in the next graph. Given the orbital velocity in these dynamics and given the fact the gravitational constant G and the orbital velocity v are constant, that is more or less independent from the distance from the galaxy center r, the only way, within the context of these Newtonian dynamics, to resolve this problem is the speculation that the total mass M contained in the sphere with radius r increases linearly with the radius.

3.- *Third part of the interpretation*. Given this speculation of the total galaxy mass increasing linearly with the radius and given the fact that the visible mass decreases rapidly with the distance from the galaxy center, it follows logically that 'dark matter', which is supposed not to interact with light, and increases with distance from the galaxy center, which implies that this dark matter is not associated to the visible matter. As a consequence of that speculation, the mass luminosity rate of the galaxy (M / L) increases dramatically, when one moves away from the galaxy center.

Graph.- Linear correlation between mass and distance to the galaxy center in Sa and Sc galaxies [10]



It is important to point out that in this interpretation, NO observations corroborate the speculations. Dark matter is not observed, precisely because it is supposed not to interact with light, as Rubin points out in a *Scientific American* article of 1983: "*All attempts to detect a halo by its visual, infrared, radio or X-ray radiation have failed. (..) In sum, the only requirement for the halo is the presence of matter in any cold, dark form that meets the* M / L *constraint.*"[11] The whole conjecture about the dark matter halo depends on the truth of the theoretical assumption that orbital velocities in spiral galaxies can be explained by Newtonian gravitational dynamics. This assumption allows for the speculation that the total galaxy mass increases linearly with distance from the center, as seen in the graph above.

[10] Original drawing by Vera Rubin in *Scientific American* (1983): 95
[11] Vera Rubin, "Dark Matter in Spiral Galaxies", en: *Scientific American* vol. 248 (1983): 98

Once the idea of a halo of exotic dark matter was published in the *Scientific American* in 1983, many cosmologists started making references to the speculation about a halo of exotic dark matter in spiral galaxies, dissociated from visible matter, as though it was a scientific fact. For example, in 1994, Kolb & Turner reproduced some flat rotation curves and affirmed that: "*Rotation curve measurements indicate that virtually all spiral galaxies have a dark, diffuse 'halo' associated with them which contributes at least 3 to 10 times the mass of the 'visible matter' (stars and the like).*"[12] In 2002, Hawking attributed to a halo of exotic dark matter, the fact that the stars at the edge of spiral galaxies like the Milky Way, NGC 3198 or NGC 9646, are maintained in their orbits and not thrown into the outer space, and presented this fact as the "*most convincing*" proof until now in favor of the existence of exotic dark matter.[13] Not all cosmologists reflected on the Newtonian assumptions of these speculations, but some did. For example, in his *The Cosmic Century* of 2006, Longair reproduced the original M31 spiral galaxy image with the flat rotation curve drawn by Rubin, and commented on the Newtonian dynamics underlying these speculations. [14] It is worthwhile quoting Longair at some length:

"*Vera Rubin and her colleagues pioneered systematic studies of the rotation curves of galaxies (...). [I]n the outer regions of galaxies, the velocity curves are generally remarkably flat,* $v_{rot} \approx$ *constant (...). The significance of this result can be appreciated from a simple* ***Newtonian calculation****. If the galaxy is taken to be spherical and the mass within the radius* r *is* M*, the circular rotational velocity at distance* r *is found by equating the inward gravitational acceleration* (GM/r^2)*, to the centripetal acceleration* $(v_{rot}{}^2/r)$*, and so* $v_{rot} = (GM/r)^{1/2}$*. Thus, if* v_{rot} *is constant, it follows that* $M \propto r$*, so that the total mass within radius* r *increases linearly with the distance from the centre. This result contrasts strongly with the variation of the surface brightness of spiral galaxies, which decrease much more rapidly with distance from the centre than as* r^{-2}*.*"[15]

Peebles too noted that Newtonian gravitational dynamics may not be applicable in the cases of galaxy clusters and spiral galaxies: "*[d]iscovering the nature of the dark matter, or explaining why the Newtonian mechanics used to infer its existence has been misapplied, has to be counted as one of the most exciting and immediate opportunities in cosmology today.*"[16] He did not follow up, however, on his own doubts:[17]

1.2.- Dark matter in galaxy clusters

The speculation about the existence of non-baryonic dark matter extends to galaxy clusters. In the next mathematical box (number 5), I explain how total galaxy cluster mass is estimated in the standard ΛCDM model, following a procedure that is based on Newtonian gravitational dynamics.

[12] Edward Kolb & Michael Turner, *The Early Universe* (1994): 17-18
[13] Stephen Hawking, *El Universo en una Cáscara de Nuez* (2005): 186
[14] Malcolm Longair, *Galaxy Formation* (2008): 66-69 y Malcolm Longair, *The Cosmic Century* (2006): 248-253
[15] Malcolm Longair, *The Cosmic Century* (2006): 248-249, bold characters are mine.
[16] James Peebles, "Dark Matter", in. *Principles of Physical Cosmology* (1993): 417
[17] James Peebles, "Dark Matter", in: *Principles of Physical Cosmology* (1993): 417-456

MATHEMATICAL BOX 2, ESTIMATING TOTAL GALAXY CLUSTER MASS IN THE CONTEXT OF NEWTONIAN GRAVITATIONAL DYNAMICS

The equation for kinetic energy is derived directly from Newton's second law of movement:[18]

$$K = \frac{1}{2}mv^2 \qquad (7)$$

and if we assume that the distribution of the velocity is isotropic, in the three directions of the system of coordinates and we assume also spherical symmetry in the galaxy cluster, we obtain:

$$K = \frac{3}{2}M\left\langle v_r{}^2 \right\rangle \qquad (8)$$

where $\langle v_r \rangle$ is the average radial velocity. Let us assume also the validity of the virial theorem, that supposes Newtonian gravitational dynamics:[19]

$$K = \frac{1}{2}\left|U_g\right| \qquad (9)$$

and we obtain the equation for potential gravitational energy, derived from Newtonian physics:[20]

$$(47)\ \left|U_g\right| = \frac{GM_1M_2}{\langle R \rangle} \qquad (10)$$

where $\langle R \rangle$ is the weighed average of the distance between objects with mass M. From equations (9) and (10), we obtain:

$$K = \frac{1}{2}GM^2 / R_{cl} \qquad (11)$$

From equations (8) and (11), we obtain (with Longair in, *Galay Formation*)

$$M = \frac{3\langle v^2 \rangle R_{cl}}{G}(^{\text{note 21}}) \qquad (12)$$

where M is the galaxy cluster mass; $\langle v \rangle$ the average rotational velocity of one galaxy; and R_{cl} the average distance between galaxies. For that reason,

$$(50)\ \langle v_{rot} \rangle = \sqrt{\frac{1}{3}\frac{GM}{\langle R_{cl} \rangle}} \qquad (13)$$

The important point to make here is that the estimate of the galaxy cluster mass[22] is based on Newtonian gravitational dynamics and, for that reason, *overestimates the total galaxy cluster mass in various orders of magnitude*, just as is the case with spiral galaxies. In estimating the

[18] See Appendix II, equations 81-91, in: John Auping, *El Origen y la Evolución del Universo* (2009): 543-545
[19] Section C1 of Appendix VIII in: John Auping, *El Origen y la Evolución del Universo* (2009): 736
[20] See Appendix II, equation 101, in: John Auping, *El Origen y la Evolución del Universo* (2009): 547
[21] See Malcolm Longair, *Galaxy Formation* (2008): 66
[22] See mathematical box 2

baryonic mass of these clusters on the basis of their luminosity,[23] and subtract this baryonic mass from the total mass,[24] as obtained from Newtonian rotational velocity, modern cosmology obtains its estimate of the total non-baryonic galaxy cluster mass, which is various times the baryonic mass. In a recent *survey* of galaxy clusters, Hans Böhringer established that the proportions of non-baryonic dark matter and baryonic visible matter are 85% and 15%, respectively and that the 15% corresponding to baryonic visible matter is distributed between stars, 2%, and gas, 13%, in big clusters; and 5% and 10%, respectively in small clusters.[25]

1.3.- The Press-Schechter theorem and dark matter

Some cosmologists have told me that they think that Press-Schechter theorem, dating from 1974, proves the existence of dark matter.[26] I donot agree. This theorem pretends to establish the number N of objects with different masses ($N_1(M_1)$, $N_2(M_2)$, $N_3(M_3)$, etc.), per volume of space (for example, Mpc^{-3}), produced by an original cloud of particles with certain initial mass (both baryonic and non-baryonic) starting to collapse because of its initial inhomogeneities or perturbations. These collapses repeat themselves at different scales, in a more or less hierarchical form, for example, on a bigger scale, the number of galaxies in a galaxy clusters and on a smaller scale, the number of stars in a galaxy.

MATHEMATICAL BOX 3. THE PRESS-SCHECHTER FUNCTION

The Press-Schechter equation gives us the number of objects N with certain mass M as a function of the critical mass M^*, related to the cause of the collapse, and of time:[27]

$$(51)\ N(M) = \frac{1}{2\sqrt{\pi}}\left(1+\frac{n}{3}\right)\frac{\bar{\rho}}{M^2}\left(\frac{M}{M^*}\right)^{(3+n)/6} \exp\left[-\left(\frac{M}{M^*}\right)^{(3+n)/3}\right] \qquad (14)$$

where "*all the time dependence of* $N(M)$ *has been absorbed into the variation of* M^* *with cosmic epoch*" and $\bar{\rho}$ is "*the mean density of the background model*",[28] n is the value of the spectral index, and the critical mass of reference M^* is defined as:[29]

$$(52)\ M^* = M_0^*\left(\frac{t}{t_0}\right)^{4/(3+n)} \qquad (15)$$

(as in Malcolm Longair's, *Galaxy Formation*), where M_0^* is the value of M^* at the present time t_0.

---

[23] See mathematical box 2
[24] See mathematical boxes 4 & 5
[25] Hans Böhringer of the Max-Planck-Institute für extraterrestrische Physik, in "Galaxy clusters as cosmological probes", lecture given at the Universidad Iberoamericana, April 16th, 2008
[26] See William Press & Paul Schechter, "Formation of galaxies and clusters of galaxies by self-similar gravitational condensation," en: *The Astrophysical* Journal, vol. 187 (1974): 425-438 y la synthesis en Malcolm Longair, "The Press-Schechter Mass Function," en *Galaxy Formation,* 2nd ed. (2008): 482-489
[27] Malcolm Longair, *Galaxy Formation* (2008):484, equation (16.25)
[28] Malcolm Longair, *Galaxy Formation* (2008):484
[29] Malcolm Longair, *Galaxy Formation* (2008):483-484, equation (16.22)

In the Press-Schechter function we find the term for the critical mass M^*, which the object must have in order to collapse, and this mass M^* is a function of the observed mass of the object at the present moment M_0^*, on the one hasnd, and the time that has passed since the original cloud started collapsing, on the other hand. The value of this critical mass at the present time (M_0^*) is determined with the laws of Newtonian gravitation analyzed before, which implies that this mass is overestimated in various orders of magnitude. The Press-Schechter function depends also of the estimate of the mean mass density $\bar{\rho}$ of the cosmological model that is used, and the validity of which is assumed. Originally, Press and Schechter used the Einstein-de Sitter model with $\Omega_0 = 1; \Omega_\Lambda = 0$, but it is also possible to use the standard model ΛCDM with $\Omega_0 = 1; \Omega_\Lambda \cong 0.7$ or any other model. Both the term for the critical mass that assumes the validity of the Newtonian gravitational dynamics, and the term for the mass density, are model-dependent, and imply previous estimates of non-baryonic dark matter. For that reason, *the Press-Schechter formalism does not prove, but rather assumes that the largest part of galaxy cluster mass is non-baryonic and is therefore compatible with that assumption.*

Besides, a critical analysis of the Press-Schechter theorem by Monaco,[30] reveals that this formalism, from the point of view of astrophysics is quite wrong but yields apparently good results: "*there is a simple, effective and wrong way to describe the cosmological mass function. Wrong of course does not refer to the results but to the whole procedure*."[31]

1.4.- The location of dark matter

In order to investigate the *location* of dark matter in clusters, advantage has been taken of the special circumstances that occur when galaxies or galaxy clusters collide and cross each other. The fact that in the case of a collision of galaxies or galaxy clusters, the stars do not collide, but the gas does, implies that the heated gas is separated from the stars. In the case of the galaxy cluster 1E0657-558 also known as the *Bullet Cluster*, Clowe and his team corroborated the fact that there are two clusters in collision, seen from aside.[32]

The clouds of hot plasma of each cluster collide and get mixed up and reduce their relative velocity, but the stars of the galaxies do not collide physically, so that the visible plasma and the galaxies are spatially separated. The separation of galaxies and plasma permits estimating the proportions of visible baryonic matter of both on the basis of their respective luminosity. By observing the effect of weak gravitational lensing —a slight distortion of the elliptic form of the galaxies—, which is more accentuated where galaxies are (with relatively little visible matter), than the plasma regions (with more visible matter), the hypothesis is corroborated that the location of the dark matter is in and around the galaxies, generating the effect of the observed gravitational *lensing*. The variations of gravitational *lensing* "*are in agreement with the galaxy*

[30] Pierluigi Monaco, "Dynamics in the Cosmological Mass Function (or, why does the Press & Schechter Function work?", in: Giuliano Giuricin & Marino Mezzetti eds., *Observational Cosmology: The Development of Galaxy Systems* (1999): 186-197

[31] *Ibidem*, pág. 187

[32] Douglas Clowe *et al*., "A direct empirical proof of the existence of dark matter", arXiv:astro-ph/0608407, reproduced in: *Astrophysical Journal Letters* (2006). Also, *idem*, "Colliding clusters shed light on dark matter," in: *Scientific American* (agosto 22, 2006)

positions and offset from the gas."[33], and the dark matter is associated to the visible matter of the galaxies. This is not a case, therefore, of 'pure' dark matter, dissociated from the visible matter. Clowe's team does not speculate about the character of the dark matter the existence of which they believe to have corroborated.[34]

Image.- Dark matter (blue) is associated to the galaxies and dissociated from the plasma (pink)

The conclusions of the analysis of the *Bullet Cluster* 1E0657-558, first realized by Clowe and his team, in 2006, then replicated by Bradac[35] and her team in the case of another *merger* of *clusters*, catalogued as MACS J0025.4-1222, and indirectly corroborated by Massey and his team, who used the observed distortion of the form of half a million galaxies to reconstruct, the distribution of the total intermediate mass that causes the distortion by lensing, are the following:

1) "*[T]he baryons follow the distribution of dark matter even on large scales*."[36]
2) There are large amounts of dark matter in these galaxy clusters. Hans Böhringer estimates the proportions of non-baryonic dark matter and baryonic visible matter at 85% and 15% (stars, 2%, and gas, 13%, in big clusters), respectively.[37]
3) The statistical methods to measure the amount of mass through weak gravitational lensing, which is at the basis of these investigations and its conclusions, is a mixture of relativistic –as far as the fact of gravitational lensing is concerned— and, in the words of Clowe, "***Newtonian gravity***",[38] or in the words of Massey, "***Newtonian***" [39] gravitational dynamics, so that this proof of the existence of dark matter rests on the validity of Newtonian gravitational dynamics in these cases.

---

[33] Douglas Clowe *et al*., "Catching a bullet: direct evidence for the existence of dark matter," arXiv:astro-ph/0611496, p. 4

[34] Dennis Zaritsky, a member of Clowe's team, admits that one does not know what this dark matter is. Quoted in "Colliding Clusters Shed Light on Dark Matter", *Scientific American* (August 22, 2006)

[35] Marusa Bradac *et al.*, "Revealing the properties of dark matter in the merging cluster MACS J0025.4-1222, en: arXiv:0806.2320

[36] Richard Massey *et al*., "Dark matter maps reveal cosmic scaffolding," in: *Nature* online, January 2008, p. 5. The dark matter is made visible by the grey contours in the first image and the grey spots in the three other ones.

[37] Hans Böhringer of the Max-Planck-Institute für extraterrestrische Physik, in "Galaxy clusters as cosmological probes", lecture given at the Universidad Iberoamericana, April 16th, 2008

[38] Douglas Clowe *et al*., "Catching a bullet: direct evidence for the existence of dark matter," arXiv:astro-ph/0611496, p. 3

[39] Richard Massey *et al*., "Probing Dark Matter and Dark Energy with Space-Based Weak Lensing", arXiv: astro-ph/0403229, p. 4, see also *idem*., "Dark matter maps reveal cosmic scaffolding, " in: *Nature* online

1.5.- Modified Newtonian Dynamics

Another possible hypothesis to explain the discrepancy between the observed and expected Newtonian dynamics of galaxy rotation velocity is the Modified Newtonian Dynamics (MOND) without dark matter as developed by Mordechai Milgrom, an astrophysicist from Israel, in different publications, beginning in 1983.[40] Milgrom maintains Newton's second law of movement, but modifies it for very slow accelerations, such as those that are common at great distances from galaxy centers.[41] The problem of this solution is the arbitrary division between 'high' and 'low' accelerations, on the one hand, and the arbitrary modification of the Newtonian dynamics, on the other hand, since it does not conform to known physical laws. Milgrom is very conscious of this fact, when formulating the following dilemma: "*Dark matter is the only explanation that astronomers can conjure up for the various mass discrepancies, if we cleave to the accepted laws of physics. But if we accept a departure from these standard laws, we might do away with dark matter.*"[42]

Mario Livio agrees with Milgrom that there are only two possible solutions to the problem, but prefers dark matter: "*There are only two ways to explain the high speeds of these clouds. (...) [E]ither Newton's law of gravitation breaks down in the circumstances prevailing in the outskirts of galaxies, or the high orbital speeds are caused by the gravitational attraction of invisible matter. (...) Astronomers have been forced to accept the second possibility: galaxies must contain large amounts of dark matter.*"[43] It is remarkable that Milgrom and Livio and many others assume that the accepted laws of physics applicable in these cases are necessarily Newtonian.

---

[40] Mordechai Milgrom, "Do Modified Newtonian Dynamics Follow from the Cold Dark Matter Paradigm?", in: *Astrophysical Journal* (may 2002)

[41] Milgrom suggests that slow accelerations produce a orbital velocity *independent of distance*: $a << a_0$ & $a_0 = 1.2 * 10^{-10}\, ms^{-2}$ & $F = a^2 / a_0 \Rightarrow\ a = \sqrt{GMa_0}\,/\,r$ which, with $a = v^2 / r$ yields $v = (GM\, a_0)^{1/4}$

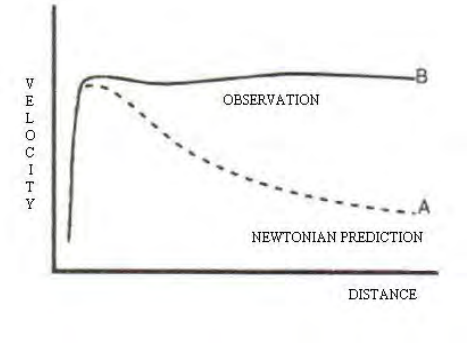
[42] Mordechai Milgrom, "Does Dark Matter Really Exist?", in: *Scientific American* (agosto de 2002), p. 44

[43] Mario Livio *The Accelerating Universe* (2000): 90

Part 2.- Cooperstock & Tieu's relativistic approach to galaxy gravitational dynamics

To understand why the assumption of Newtonian gravitational dynamics is not necessary, we have to reformulate the essence of the problem:

Graph.- The observed and expected Newtonian dynamics of galaxy rotation velocity



A few cosmologists grasped the opportunity referred to by Longair and Peebles, as to the Newtonian dynamics at the basis of the dark matter speculation, notably two teams from Canada, Brownstein and Moffat, on the one hand, and Cooperstock and Tieu, on the other hand, who offered, one team independently of the other, twelve years after Peeble's proposal, *an orthodox solution, along the lines of Einstein's general relativity, that makes the speculations about dark matter superfluous.* According to Cooperstock and Tieu, galactic dynamics present a non-linear, relativistic problem. Eddington had mentioned this non-linearity for a system that is variable in time, and the authors extend it to non-linear, but stationary (non-time dependent) problems, as in galactic dynamics:

"*In dismissing general relativity in favor of Newtonian gravitational theory for the study of galactic dynamics, insufficient attention has been paid to the fact that the stars that compose the galaxies are essentially in motion under gravity alone ('gravitationally bound'). It has been known since the time of Eddington that the gravitationally bound problem in general relativity is an intrinsically non-linear problem even when the conditions are such tat the field is weak and the motions are non-relativistic, at least in the time-dependent case. Most significantly, we found that under these conditions, the general relativistic analysis of the problem is also non-linear for the stationary (non-time-dependent) case at hand. Thus the intrinsically linear Newtonian-based approach used to this point has been inadequate for the description of galactic dynamics (...). We ... demonstrate that via general relativity, the generating potentials producing the observed flattened galactic rotation curves are necessarily linked to the mass density distributions of the flattened disks [of ordinary baryonic matter], obviating any necessity for dark matter halos in the total galactic composition.*"[44]

Cooperstock, colaborating first with Tieu and then with Carrick, has analyzed a total of 7 spiral galaxies from the relativistic point of view:

[44] Cooperstock & Tieu, General Relativity Resolves Galactic Rotation Without Exotic Dark Matter (2005): arXiv:astro-ph/0507619, ps. 2-3

1) In their first publication of 2005, the authors analyzed four spiral galaxies (the *Via Lactea*, and NGC 3031, NGC 3198, and NGC 7331)[45]
2) In December 2010 they added another three spiral galaxies proving the same point (NGC 2841, NGC 2903 and NGC 5033)[46]

Cooperstock, Tieu and Carrick conceive the sprial galaxies as systems that are analogous to "*fluids rotating uniformly without pressure and symmetric around the axis of rotation*,"[47] and explained the rotational dynamics by the gravitational attraction exercised by baryonic matter, within the known form of the visible disk, *in relativistic gravitational dynamics* (see the next mathematical box).

MATHEMATICAL BOX 4. SPIRAL GALAXY MASS IN RELATIVISTIC, NON-LINEAR GRAVITATIONAL DYNAMICS ACCORDING TO COOPERSTOCK

Cooperstock & Tieu start from the line element of anobject in 'free fall' in general relativity, adapted to the polar, cylindrical coordinates r y z:

$$ds^2 = -e^{v-w}(u\,dz^2 + dr^2) - r^2 e^{-w} d\phi + e^{w}(c\,dt - N\,d\phi)^2 \qquad (16)$$

where u, v, w, and N are coefficients whose value is a function of the coordinates r and z. For various reasons, explained by the authors[48] one may simplify this equation equating $u = 1$ and $w = 0$:

$$ds^2 = -e^{v}(dz^2 + dr^2) - r^2 d\phi + (c\,dt - N\,d\phi)^2 \qquad (17)$$

We obtain the relation between angular velocity ω, and tangential velocity V and the coefficient N (using $\bar{\phi} = \phi + \omega(r,z)t$):[49]

$$\omega = \frac{N c}{r^2} \qquad (18)$$

and (14) $V = \omega r \qquad (19)$

so that, by (18) and (19):

$$V = \frac{N c}{r} \qquad (20)$$

The authors use Einstein's field equations for N and ρ in a weak field with a cloud of particles in rotational motion, not subject to pressure neither to friction:[50]

[45] Fred Cooperstock & Steven Tieu, General Relativity Resolves Galactic Rotation Without Exotic Dark Matter (2005): arXiv:astro-ph/0507619

[46] J. D. Carrick and F. I. Cooperstock, General relativistic dynamics applied to the rotation curves of galaxias, arXiv:1101.3224, December 2010

[47] Fred Cooperstock & Steven Tieu, "General Relativity Resolves Galactic Rotation Without Exotic Dark Matter" (2005): arXiv:astro-ph/0507619, p. 4

[48] Fred Cooperstock & Steven Tieu, "Galactic Dynamics via General Relativity", en: *International Journal of Modern Physics* vol. 22 (2007): 4-5

[49] Fred Cooperstock & Steven Tieu, "Galactic Dynamics via General Relativity", in *International Journal of Modern Physics* vol. 22 (2007): 4

[50] Fred Cooperstock & Steven Tieu, "Galactic Dynamics via General Relativity", in: *International Journal of Modern Physics* vol. 22 (2007): 5

$$N_{rr} + N_{zz} - \frac{N_r}{r} = 0 \qquad (21)$$

$$\text{and} \qquad \frac{N_r{}^2 + N_z{}^2}{r^2} = \frac{8\pi G\rho}{c^2} \qquad (22)$$

Equation (21) can be represented as a function of the gravitational potential ϕ for rotating galaxies:

$$\nabla^2\phi = 0 \qquad (23)$$

where the zero value is due to the absence of pressure and friction in a system of particles in rotational motion. If there were no rotational motion, the system would need pressure (a non-zero value) to be stable, as in the Poisson equation of Newtonian gravity for weak fields:

$$\nabla^2\phi = 4\pi G\rho \qquad (24)$$

In a way analogous to the derivation of the Newtonian gravitational field and potential,[51] Cooperstock and Tieu obtain the gravitational potential of a system of particles in rotational motion not subject to pressure, neither to friction:

$$\phi = \int \frac{N}{r} dr \quad \Rightarrow \quad \frac{\partial\phi}{\partial r} = \frac{\partial}{\partial r}\int \frac{N}{r} dr = \frac{N}{r} \qquad (25)$$

From equations (20) and (25), we obtain the rotational or tangential velocity:

$$V = c\frac{N}{r} = c\frac{\partial\phi}{\partial r} \qquad (26)$$

In polar cylindrical coordinates, the solution to equation (23) is: [52]

$$\phi = Ce^{-k|z|}J_0(kr) \qquad (27)$$

where J_0 is the Bessel function $J_m(kr)$ of cero order ($m = 0$) and C is an arbitrary constant. Given the fact that equation (18) is linear, we can rewrite equation (27) as a linear summary:

$$\phi = \sum_n C_n e^{-k_n|z|}J_0(k_n r) \qquad (28)$$

From equations (26) and (27), we obtain:

$$V = c\frac{\partial\phi}{\partial r} = -c\sum_n C_n k_n e^{-k_n|z|}J_1(k_n r) \qquad (29)$$

and from (26) and (29), we obtain:

$$N = \frac{Vr}{c} = -\sum_n C_n k_n r\, e^{-k_n|z|}J_1(k_n r) \qquad (30)$$

[51] For the derivation of the Newtonian gravitational field and potential, see Appendixes II & VI B of John Auping, *Origen y Evolución del Universo* (2009). The Poisson equation is number (239) on page 668

[52] Fred Cooperstock & Steven Tieu, "Galactic Dynamics via General Relativity", in: *International Journal of Modern Physics* vol. 22 (2007): 9

By solving $\sum_n C_n k_n$ from $n=1$ to $n=10$, we obtain the theoretical rotational velocity curves, which are perfectly corroborated by observations, with the Bessel function of order one:

$$V(r,z) = -c\sum_{n=1}^{10} C_n k_n e^{-k_n|z|} J_1(k_n r) \quad (31)$$

and $$N(r,z) = -\sum_{n=1}^{10} C_n k_n r\, e^{-k_n|z|} J_1(k_n r) \quad (32)$$

and from (31) and (32), and taking into account that $c = 3*10^8 m/s$, we obtain:[53]

$$V(r,z) = \frac{3*10^8}{r} N(r,z) \quad (33)$$

Each galaxy is different, and has its proper coefficients C_n and k_n. Cooperstock and Tieu attached the respective values of these coefficients to their article for four galaxies, among them the Milky Way, for $n=1$ to $n=10$.

What many astrophysicists and cosmologists attribute to a halo of cold dark matter in the context of Newtonian gravitational dynamics is explained by Cooperstock and Tieu with ordinary *baryonic matter in the context of relativistic gravitational dynamics*. This method yields the following results: "*Most significantly, our correlation of the flat velocity curve is achieved with disk mass of an order of magnitude smaller than the envisaged halo mass of exotic dark matter.*"[54]

Their hypothesis is corroborated: "*We have seen that the non-linearity for the computation of density inherent in the Einstein field equations for a stationary axially-symmetric pressure-free mass distribution, even in the case of weak fields, leads to correct galactic velocity curves as opposed to the incorrect curves that had been derived on the basis of Newtonian gravitational theory.*"[55] As a mater of fact, the observations corroborate the predictions of the theoretical model, as can be appreciated in the following graph by Cooperstock y Tieu: [56]

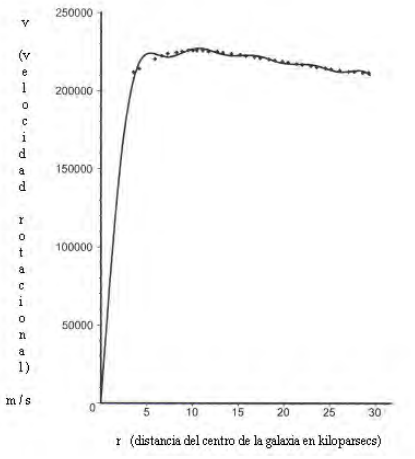
[53] *Ibidem,* Appendix, p. 30

[54] Fred Cooperstock & Steven Tieu, "General Relativity Resolves Galactic Rotation Without Exotic Dark Matter", arXiv:astro-ph/0507619 (2005): 11

[55] Fred Cooperstock & Steven Tieu, "Galactic Dynamics via General Relativity", in: *International Journal of Modern Physics A* vol. 22 (2007): 29

[56] Fred Cooperstock & Steven Tieu, "General Relativity Resolves Galactic Rotation Without Exotic Dark Matter", arXiv:astro-ph/0507619 (2005) and "Galactic Dynamics via General Relativity", in: *International Journal of Modern Physics A* vol. 22 (2007): 7.

Graph.- The rotational velocity curve for the *Milky Way* as predicted by the relativistic theory of Cooperstock y Tieu (the curve) is corroborated by the observations (the points)

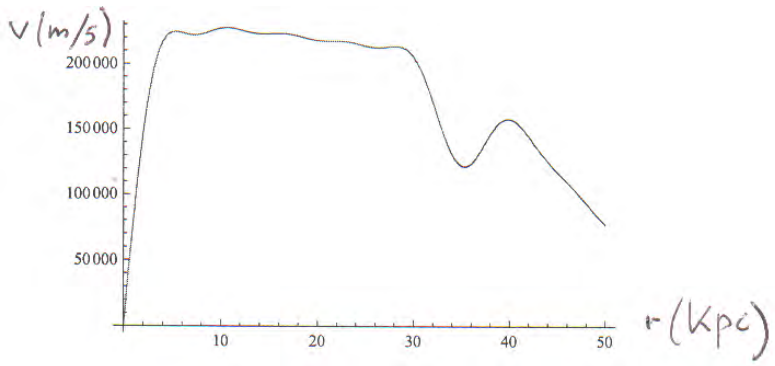


It is essential in science that the experiments set up by the authors can be replicated by colleagues. Some cosmologists expressed doubts to me as to the replicability of the proofs offered by Cooperstock and Tieu. With Wolfram's program *Mathematica*, version 6, Alfredo Sandoval and I were able to reproduce exactly the same flat rotational velocity curves. The values and the equations of Mathematical Box 4 allowed us to reproduce with Wolfram's program *Mathematica*, the same rotational velocity curves as published by Cooperstock and Tieu We discovered, however, that variations in the fourth or fifth or sixth decimal figure of the value of the coefficients C_n and k_n may affect the results in a non-trivial way. This means we can not use figures that are rounded up to the third or fourth decimal. The following graph replicates the results by Cooperstock and Tieu, for the Milky Way, in the context of relativistic gravitational dynamics. The only difference with the previous graph by Cooperstock and Tieu is that these authors give the results up to a distance of 30 kilo parsecs, and Sandoval and I, up to a distance of 50 kilo parsecs.

Cooperstock and Tieu comment their findings: "*The scientific method has been most successful when directed by 'Ockham's razor', that new elements should not be introduced into a theory unless absolutely necessary. If it should turn out to be the case that the observations of astronomy can ultimately be explained without the addition of new exotic dark matter, this would be of considerable significance.*"[57]

[57] Fred Cooperstock & Steven Tieu, "Galactic Dynamics via General Relativity", in *International Journal of Modern Physics A* vol. 22 (2007) y arXiv:astro-ph/0610370, p.30

Graph.- The rotational velocity curve for the Milky Way by Cooperstock y Tieu is replicated with Wolfram's program *Mathematica* by Sandoval and the author



The work of Cooperstock and Tieu has generated much interest and also some public criticism, from Korzynski,[58] Vogt and Letelier,[59] and Garfinkle.[60] Cooperstock and Tieu responded adequately to their critics.[61]

Other astropysicists have expressed the criticism that *even with relativistic gravitational dynamics, the baryonic mass in the spiral galaxies is not enough to explain their rotational velocity*. One of them wrote to me in an e-mail of December 2010, that "*the masses of the galaxies Cooperstock and Tieu find are greatly in excess of any reasonable estimate of the baryonic mass.*" What these astrophysicists sustain is that even in the case of relativistic gravitational dynamics only the sum of baryonic *and dark matter* in spiral galaxies explains the rotational velocity curves of stars belonging to the galaxy. They refer to Stephen Kent's estimates of spiral galaxy mass, because Cooperstock & Tieu themselves make that comparison. Kent has three articles on this topic, published in *The Astronomical Journal* in 1986, 1987 and 1988, titled *Dark matter in spiral galaxies I*; *II*; and *III*, respectively.[62]

I do not think that Kent's estimates validate the missing mass hypothesis, as I will now show. First, there is no indication in Kent's figures of a 1/6 baryon/total mass ratio as Cooperstock himself and his critics assert. Kent does not use the term 'baryonic mass', but refers to 'stellar mass', being the sum of the stellar 'bulge' mass M_B and stellar 'disk' mass M_D that he obtains by means of estimates of the mass/luminosity ratio M/L (the luminosity being the surface brightness of stars) for bulges and disks. He derives the total mass estimate M_{tot} by means of Newtonian equations that establish a causal relationship between mass and rotational velocity at

[58] Nikolaj Korzynski, "Singular disk of mater in the Cooperstock-Tieu galaxy model," arXiv:astro-ph/0508377
[59] Daniel Vogt & Patricio Letelier, "Presence of exotic matter in the Cooperstock and Tieu galaxy model," arXiv:astro-ph/0510750
[60] David Garfinkle, "The need for dark matter in galaxies", arXiv:gr-qc/051182
[61] Fred Cooperstock & Steven Tieu, "Perspectives on Galactic Dynamics via General Relativity," arXiv:astro-ph/0512048 y "Galactic Dynamics via General Relativity", in: *International Journal of Modern Physics A* vol. 22 (2007): ps. 17-28
[62] Stephen M. Kent, "Dark Matter in Spiral Galaxies. I. Galaxias with Optical Rotation Curves", *The Asytrophysical Journal,* vol. 9 (June 1986): 1301-1327; "Dark Matter in Spiral Galaxies. II. Galaxies with H1 Rotation Curves", *The Asytrophysical Journal,* vol. 9 (April 1987): 816-832; "Dark Matter in Spiral Galaxies. III. The Sa Galaxies, *The Asytrophysical Journal,* vol. 9 (August 1988): 514-527

any chosen radius, where velocity is the observed part (derived from blue and redshifts), radius the chosen part, and mass the inferred part. The stellar/total mass ratio, for 7 of the 16 galaxies with complete data sets, is $M_{B+D} / M_{tot} = 0.32$, that is a ratio of 1/3, not 1/6.[63]

Of the three galaxies chosen by Cooperstock & Tieu for analysis in their first article (NGC 3031; NGC 3198; and NGC 7331), Kent has complete data sets only for two of them, NGC 3031 and NGC 7331, since he has no bulge mass estimate fort NGC 3198. In the case of NGC 3031 and NGC 7331, the stellar/total mass ratios are 1/2 and 1/4, respectively. These ratios, however, have to be corrected, because Kent neither includes estimates of interstellar gas, nor of black holes, but "*the contribution ... from the stellar component alone*".[64]. In galaxy clusters the gas mass is 85% of the baryonic mass, and the stellar mass, 15%. In galaxies, it is 15% and 85%. The interstellar gas mass, for example in the *Via Lactea*, is 15% of the total baryonic mass.[65]. The additional gas mass changes the Kent estimate of baryon mass for NGC 3031 and NGC 7331 to $8.47*10^{10}$ and $14.24*10^{10}$ solar masses, implying 1/1½ and 1/3 baryon/total mass ratios, respectively.

Now we have to add the black hole masses M_{BH}, which is originally baryonic dark matter. The black hole mass is *not* included in Kent's M/L ratios because in both galaxies, he has identical M/L ratios for bulges and disks -3.76 and 4.04, respectively—, as he explains: "*for NGC 3031 ... and NGC 7331, the bulge M/L ratio in the full solution was very poorly constrained and so it was kept fixed equal to the disk M/L ratio*".[66] The M/L ratios would have been different for bulges and disks if the black hole had been included in the bulge M/L ratio. In NG 3031, the black hole at the center of the galaxy has a negligible mass of $M_{BH} \approx 10^8$ solar masses[67], but the authors estimate the total stellar mass to be $M_{B+D} = 7.7*10^{10}$ solar masses, which increases the baryon mass estimate to 8.97*10^10. The NGC 7331 black hole may be as much as 10^9 solar masses[68] which increases the baryon mass from 14.24 to 14.34*10^10.

The following table gives the corrected Kent baryon mass estimates, compared to those found by Cooperstock, and shows *the baryon masses in both cases are in fact almost identical*.

Galaxy	Cooperstock	Kent corrected	Kent/Cooperstock
NGC 3031	10.9 *10^10	8.97*10^10	0.82
NGC 3198	10.1*10^10	/	/
NGC 7331	26.0*10^10	14.34*10^10	0.55

[63] Stephen Kent, "Dark Matter in Spiral Galaxies. II. Galaxies with H1 Rotation Curves", *The Asytrophysical Journal,* vol. 9 (April 1987): 827

[64] Stephen Kent, ""Dark Matter in Spiral Galaxies. I. Galaxias with Optical Rotation Curves", *The Asytrophysical Journal,* vol. 9 (June 1986): 1301

[65] Katia Ferrière, "The interstellar environment of our galaxy", arXiv:astro-ph/0106359 (June 2001): 1-56

[66] Stephen Kent, "Dark Matter in Spiral Galaxies. II. Galaxies with H1 Rotation Curves", *The Asytrophysical Journal,* vol. 9 (April 1987): 826

[67] Rohlfs & Kreitschmann, "A Two component mass model for M81/NGC3031"*Astronomy & Astrophysicis*, (1980):175-182

[68] Silchenko, "Chemically decoupled nuclei in the spiral galaxies NGC 4216 and 4501", *Astronomical Journal* (1999):186-196

The table reveals that in the case of NGC 7331 there is still is some apparent missing mass, but not much. The 'missing mass' may be due to many uncertainties, among other things:
1) the uncertain estimates of the M/L ratio;
2) "*relationship between luminous and dark matter shows significant variation among galaxies*";
3) "*the relative amounts of dark and luminous matter in a galaxy are still not well known*"[69]; and
4) the "*optical rotation curves usually do not place strong constraints on the amount of dark matter in these galaxies. Indeed, in agreement with Kalnajs, 1983, some rotation curves are fit well without the need to assume the existence of any dark halo*".[70]
5) The mass of supermassive black holes at the galaxy center may have been underestimated.[71]
There are other ways in which black hole dynamics falsify the dark matter hypothesis. Though I donot share the MOND theory, their adherents have convincingly falsified the dark matter hypothesis in galaxies,[72] adding this proof to the one offered by Cooperstock ansd his colaborators.

In the case of NGC 2841, NGC 2903 and NGC 5033, Cooperstock again employed the solution of the Einstein field equations of general relativity, as in his earlier studies, and the fits to the data appeared again to be very precise. The known data from galactic rotation curves can be accommodated with "*at most relatively little extra [baryonic] matter*" when the analysis is performed with Einstein's as opposed to Newton's gravity. This 'little extra baryonic matter' may be "*due to dead stars, planets, neutron stars [black holes] and other normal non-luminous baryonic matter debris*."[73]

[69] Stephen Kent, "Dark Matter in Spiral Galaxies. II. Galaxies with H1 Rotation Curves", *The Asytrophysical Journal,* vol. 9 (April 1987): 816
[70] Stephen Kent, ""Dark Matter in Spiral Galaxies. I. Galaxias with Optical Rotation Curves", *The Asytrophysical Journal,* vol. 9 (June 1986): 1301, 1326
[71] If the black hole at the center of NGC 3031 would have a mass equivalent to the heaviest one known today ($\approx 1.8*10^{10}$ solar masses), then Cooperstock's and Kent's NGC 3031 baryon estimates would be exactly equal.
[72] Karl Gebhardt *et al*., "A relationship between nuclear black hole mass and galaxy velocity dispersion", Astrophysical Journal Letters 539 (2000): 75 ss.. See also several recent contributions by Pavel Kroupa.
[73] J. D. Carrick and F. I. Cooperstock, General relativistic dynamics applied to the rotation curves of galaxias, arXiv:1101.3224, December 2010, ps. 3,10

Part 3.- Brownstein & Moffat's relativistic approach to gravitational galaxy dynamics

Independently of Cooperstock and Tieu, another Canadian team, consisting of Brownstein and Moffat, designed a relativistic gravitational model, called *Modified Gravity* (MOG), in which they modify the Newtonian laws of acceleration on the basis of the theory of general relativity.[74] They corroborated their theory with the data of more than 160 galaxies.[75] The following mathematical box synthesizes the essence of this new relativistic theory.

MATHEMATICAL BOX 5. THE MASS OF SPIRAL GALAXIES IN RELATIVISTIC GRAVITATIONAL DYNAMICS

By adding the relativistic acceleration to the Newtonian one, one finds tha relativistoc equation of gravitational force, as I explain in my book on cosmology:[76]

$$F_{EINSTEIN} = m(\bar{a}_{NEWTON} + \bar{a}_{EINSTEIN}) = m\{\frac{d^2x^\eta}{(ds)^2} + \Gamma^\eta_{\mu\nu}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds}\} \qquad (34)$$

Brownstein and Moffat start with an analogous, relativistic modification of the Newtonian acceleration law:

$$a(r) = -\frac{G_\infty M}{r^2} + G_0\sqrt{M_0 M}\left\{\frac{e^{-r/r_0}}{r^2}(1+\frac{r}{r_0})\right\} \qquad (35)$$

and given the following effective gravitational constant:

$$G_\infty = G_0\left(1+\sqrt{\frac{M_0}{M}}\right) \qquad (36)$$

and by combining (35) y 36), they obtain:

$$a(r) = -\frac{G_0 M}{r^2}[1+\sqrt{\frac{M_0}{M}}\left\{1-e^{-r/r_0}(1+\frac{r}{r_0})\right\}] \qquad (37)$$

As Brownstein explained to me in an e-mail, a year ago, the terms M_0 and r_0 do NOT represent a variable mass and a variable radius, but are parameters the values of which are constant. In the case of high surface brightness galaxies (HSB), the values of these constants are:

$$M_0 = 9.60*10^{11} M_{SOL} \qquad (38)$$

$$\text{and} \quad r_0 = 13.92\ kpc = 4.30*10^{20} m \qquad (39)$$

In the case of very low surface brightness (LSB) or dwarf galaxies, the mass and radius have the following values:

$$M_0 = 2.40*10^{11} M_{SOL} \qquad (40)$$

[74] Joel Brownstein & John Moffat, "Galaxy Rotation Curves Without Non-Baryonic Dark Matter", arXiv:astro-ph/0506370 (2005)

[75] Joel Brownstein & John Moffat, "Galaxy Rotation Curves Without Non-Baryonic Dark Matter", arXiv:astro-ph/0506370 (2005): 18-28

[76] Equation (170 B) of Appendix VI B, in: John Auping, *The Origen and Evolution of the Universe* (2009): 660

$$r_0 = 6.96\,kpc = 2.15*10^{20}\,m \qquad (41)$$

Given the fact that according to Newton:[77]

$$a(r) = \frac{v^2}{r} \qquad (42)$$

we obtain, combining equations (37) and (42), the law of modified velocity:

$$v(r) = \sqrt{\frac{G_0 M}{r}}\left[1 + \sqrt{\frac{M_0}{M}}\left\{1 - e^{-r/r_0}\left(1 + \frac{r}{r_0}\right)\right\}\right]^{1/2} \qquad (43)$$

In the case of a symmetric galaxy, the mass density of which contains an interior *core* at a distance $r = r_c$, the acceleration of equation (37) is transformed in equation (44) for HSB galaxies and in (45) for LSB and dwarf galaxies LSB, respectively:

$$\text{HSB:}\quad a(r) = -\frac{G_0 M\left(\frac{r}{r + r_c}\right)^3}{r^2}\left[1 + \sqrt{\frac{M_0}{M}}\left\{1 - e^{-r/r_0}\left(1 + \frac{r}{r_0}\right)\right\}\right] \qquad (44)$$

$$\text{LSB:}\quad a(r) = -\frac{G_0 M\left(\frac{r}{r + r_c}\right)^6}{r^2}\left[1 + \sqrt{\frac{M_0}{M}}\left\{1 - e^{-r/r_0}\left(1 + \frac{r}{r_0}\right)\right\}\right] \qquad (45)$$

In these cases, the rotational velocity for HSB galaxies is transformed in equation (46) and for LSB and dwarf galaxies in (47), respectively:

$$\text{HSB:}\quad v(r) = \sqrt{\frac{G_0 M}{r}}\left(\frac{r}{r + r_c}\right)^{3/2}\left[1 + \sqrt{\frac{M_0}{M}}\left\{1 - e^{-r/r_0}\left(1 + \frac{r}{r_0}\right)\right\}\right]^{1/2} \qquad (46)$$

$$\text{LSB:}\quad v(r) = \sqrt{\frac{G_0 M}{r}}\left(\frac{r}{r + r_c}\right)^{3}\left[1 + \sqrt{\frac{M_0}{M}}\left\{1 - e^{-r/r_0}\left(1 + \frac{r}{r_0}\right)\right\}\right]^{1/2} \qquad (47)$$

These equations of the acceleration and rotational velocity differ from the classical, Newton rotational ones:

$$a_{NEWTON} = -\frac{G_0\left(\frac{r}{r + r_c}\right)^{3\beta}}{r^2} \qquad (48)$$

with $\beta = 1$ for HSB and $\beta = 2$ for LSB and dwarf galaxies.

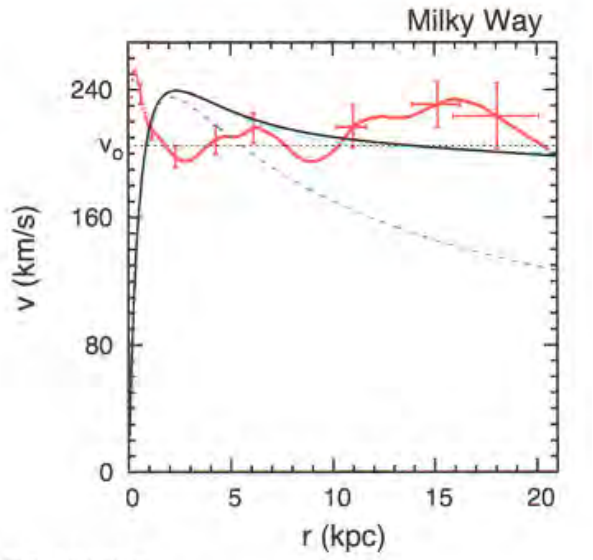
$$v_{NEWTON} \approx \sqrt{\frac{G_\infty M}{r}} \text{ (where } r \to \infty) \qquad (49)$$

Even though Cooperstock & Tieu and Brownstein & Moffat, use different relativistic models, the results are identical, as can be appreciated comparing, for example, the rotational velocity

[77] See mathematical box 1

curves of the Milky Way, produced by both teams. In the following graph, I reproduce the rotational velocity curve of the Milky Way, produced by Brownstein and Moffat,[78] that can be compared with the one produced by Cooperstock y Tieu, reproduced above..

Graph.- The rotational velocity curve of the Milky Way according to the relativistic theory MOG



EXPLANATION:
Red points: observations of rotational velocity
Solid black curve: curve obtained by relativistic MOG (MSTG) theory
Dashed blue line: predictions of Newtonian theory
Green curve (behind the black curve): prediction of MOND

In synthesis, the gravitational dynamics of spiral galaxies are well explained by Einstein's theory of general relativity, without any necessity to introduce speculations about a halo of non baryonic dark matter

Image.- The dynamics of spiral galaxies is explained by Einstein's general relativity [79]

[78] Joel Brownstein & John Moffat, "Galaxy Rotation Curves Without Non-Baryonic Dark Matter", arXiv:astro-ph/0506370, p. 29

[79] The spiral galaxy NGC 6946. Photo by John Duncan, *Astronomía* (2007): 223

Part 4.- Relativistic dynamics of galaxy clusters without cold dark matter

4.1.- Cooperstock & Tieu

¿Can we extend this analysis of galaxy rotational velocity to galaxy clusters? Cooperstock & Tieu have shown that it is indeed possible[80]: "*For the dynamics of clusters of galaxies, the virial theorem is used. This is based on Newtonian gravity theory. It would be of interest to introduce a general relativistic virial theorem for comparison. It is only after possible effects of general relativity are explored that we can be confident about the viability or non-viability of exotic dark matter in nature.* "[81]

As a matter of fact, in 2008, Cooperstock and Tieu applied general relativity to the gravitational dynamics of galaxy clusters, and corroborated their hypothesis,[82] especially their hypothesis about total cluster mass and the rotational velocity of galaxies.[83] The following mathematical box synthesizes their main argument, based on a relativistic model of a weak gravitational field *constituted by many bodies that suffer mutual gravitational attraction but no friction or pressure.*

MATHEMATICAL BOX 6. TOTAL MASS ESTIMATE OF GALAXY CLUSTERS IN A RELATIVISTIC GRAVITATIONAL MODEL

Cooperstock and Tieu start with Schwarzschild's solution to Einstein's equations, that uses a metric of spherical coordinates for a spherical mass M,[84] the same that one uses to derive the perihelion rotation of Mercury in a plane:[85]

$$ds^2 = \left(\frac{1}{1 - \frac{2GM}{c^2 r}} \right) dr^2 + r^2 (d\theta^2 + sen^2\theta \, d\phi^2) - \left(1 - \frac{2GM}{c^2 r} \right) c^2 dt^2 \qquad (50)$$

The four terms between parentheses constitute the metric coefficients, that together determine Schwarzschild's metric tensor in fourdimensional space-time, where the

[80] Fred Cooperstock & Steven Tieu, "Perspectives on Galactic Dynamics via General Relativity," arXiv:astro-ph/0512048 (2005) and Fred Cooperstock, "Clusters of Galaxies", in: *General Relativistic Dynamics* (2009): 135-159

[81] Fred Cooperstock & Steven Tieu, "Perspectives on Galactic Dynamics via General Relativity," arXiv:astro-ph/0512048, p. 3. For the virial theorem, see Appendix VIII, Section C 1, in: John Auping, *El Origen y la Evolución del Universo* (2009), p. 736

[82] Fred Cooperstock & Steven Tieu, "General relativistic velocity", in: *Modern Physics Letters A* vol. 23 (2008): 1745-1755 and Fred Cooperstock, "Clusters of Galaxies", in: *General Relativistic Dynamics* (2009): cap. 10

[83] John Moffat, "Scalar-Tensor-Vector Gravity Theory", arXiv:gr-qc/0506021; "A Modified Gravity and Its Consequences for the Solar System, Astrophysics and Cosmology," arXiv:gr-qc/0608074; y Joel Brownstein & John Moffat, "Galaxy Cluster Masses Without Non-Baryonic Dark Matter," arXiv:astro-ph/0507222; and *Monthly Notices of the Royal Astronomical Society* (2005): 1-16

[84] See equations (381) and (382) of appendix VI B of John Auping, *Origen y Evolución del Universo* (2009): 692. Cooperstock and Tieu invert the signs and simplify the equation, omitting the constants G and c, see Fred Cooperstock & Steven Tieu, "General relativistic velocity: the alternative to dark matter", in: *Modern Physics Letters A* vol. 23 (2008): 1746, equations (1) and (2)

[85] See equation (382) of Appendix VI B and equation (4) of Appendix VI C of John Auping, *Origen y Evolución del Universo* (2009): 692, 698

mass M is not small. In the normalized version of Cooperstock and Tieu, the units are selected so as to make $c = G = 1$ and the signs of the metric are inverted:[86]

$$ds^2 = -\left(1-\frac{2m}{r}\right)^{-1} dr^2 - (r^2)d\theta^2 - (r^2 sen^2\theta)\, d\phi^2 + \left(1-\frac{2m}{r}\right)dt^2 \qquad (51)$$

A big difference between this procedure and the perihelion analysis of Mercury is that we do not make the simplifying assumption, justifiable in the solar system, that the proper time $d\tau$ of the observed mass and the time dt of the observer are one and the same. Normally, in the case of strong gravitational fields, this difference is considered to be important. But, Cooperstock and Tieu show that also in the case of a weak gravitational field, the difference between the proper time $d\tau$ of the observed mass and the time dt of the observer is crucial. The transformation of the coordinates of the observer (r and t) into the *'co-moving'* coordinates of the observed object with its proper time (R and τ) is the following:

$$\tau = t + \int \frac{\sqrt{\frac{2m}{r}}}{1-\frac{2m}{r}} dr \qquad (52)$$

$$R = t + \int \frac{1}{\sqrt{\frac{2m}{r}}(1-\frac{2m}{r})} \qquad (53)$$

$$\text{and} \qquad r = \left(\frac{3}{2}(R-r)\right)^{2/3} (2m)^{1/3} \qquad (54)$$

which gives us the following transformed Schwarzschild metric, that Cooperstock took from Landau & Lifshitz's *The Classical Theory of Fields*, and that depends on the proper time of the massive object:[87]

$$ds^2 = d\tau^2 - \frac{dR^2}{\left(\frac{3}{2(2m)}(R-\tau)\right)^{2/3}} - \left(\frac{3}{2}(R-\tau)\right)^{4/3} (2m)^{2/3}(d\theta^2 + sen^2\theta\, d\varphi^2) \qquad (55)$$

In the case that the value of τ comes close to the value of R, we are in a strong gravitational field and the singularity of a black hole arises where $R = \tau$. But Cooperstock and Tieu are interested in the case of a *weak* gravitational field, where $R >> \tau$ for all R, implying that $r >> 2m$ for all r and the coordinates (r,t). The

[86] Fred Cooperstock & Steven Tieu, "General relativistic velocity: the alternative to dark matter", in: *Modern Physics Letters A* vol. 23 (2008): 1746, equations (1) and (2) and note 6 ($c = G = 1$). This is equation (100.2) of L. Landau & E. Lifshitz, *The Classical Theory of Fields,* 4ª ed. revisad (2002):321, if one takes into account that r_g (the 'gravitational radius') in Landau y Lifshitz is the mass m in Cooperstock and Tieu.

[87] This is equation (102.3) of L. Landau & E. Lifshitz, *The Classical Theory of Fields,* 4ª revised ed. (2002):332, if one takes into account that r_g (the gravitational radius) in Landau and Lifshitz is the mass m in Cooperstock and Tieu, and that Landau and Lifshitz normalize only half way ($G = 1$, but $c \neq 1$).

radial velocity, measured by the external observer is:

$$v_{rad} = \frac{dr}{dt} = -\left(1 - \frac{2m}{r}\right)\left(\sqrt{\frac{2m}{r}}\right) \qquad (56)$$

The radial velocity in the proper time of the observed moving object is:

$$v_{rad} = \frac{dR}{d\tau} = -\sqrt{\frac{-g_{11}}{g_{00}}}\frac{dr}{dt} = -\sqrt{\frac{1}{(1-2m/r)^2}}\left(1 - \frac{2m}{r}\right)\sqrt{\frac{2m}{r}} \cong -\sqrt{\frac{2m}{r}} \ (^{\text{nota 88}}) \qquad (57)$$

In a weak gravitational field, the radial velocity measured in the proper time of the co-moving object is equal to the radial velocity measured in the time of the terrestrial observer, because the mass m of the field is so reduced that the factor $\left(1 - \frac{2m}{r}\right) \approx 1 - 0 = 1$, as Cooperstock and Tieu explain: "*the local measures, both proper and external, of the radial velocity are approximately equal in the value of* $-\sqrt{2m/r}$*.*"[89] However, this is only true in the case that almost all the mass of the systrem is concentrated in the centre of mass, as for example in the solar system. So far, the weak gravitational field is originated by one massive object. But things get complicated, when we focus on the collapse of a cloud of particles*, where each particle contributes to the total mass and field*. It is in this case, that the radial velocity as measured in the proper time of the co-moving object, even in the case of non-relativistic velocities, starts differing considerably from the time of the external, terrestrial observer. Parting from the geodesic equation in general relativity, for a cloud of dust particles, taken from the classic work of Landau and Lifshitz,[90] Cooperstock and Tieu obtain the following geodesic equation for dust particles or objects, as measured by an external (terrestrial) observer:

$$ds^2 = d\tau^2 - e^{\lambda(\tau,R)}dR^2 - r^2(\tau,R)(d\theta^2 + sen^2\theta\, d\varphi^2) \qquad (58)$$

In this case, "*a freely falling dust particle maintains constant space coordinates for all time*," and the exact solution of the four non-trivial Einstein field equations that apply in this case, "*assumes a surprisingly simple form*" in the form of the following two equations:[91]

---

[88] The Schwarzschild metric in Cooperstock and Tieu is $g_{00} = (1-2m)/r$ and $g_{11} = -(1/(1-2m/r))$, the difference with the Schwarzschild metric in John Auping, *Origen y Evolución del Universo* (2009), Appendix VI B, equation 382, p. 692, is that g_{00} in Cooperstock and Tieu is my $-g_{44}$ ($g_{44} = -(1-2GM/c^2 r)$), and g_{11} in Cooperstock and Tieu is my $-g_{11}$ ($g_{11} = 1/(1-2GM/c^2 r)$), with C. & T. normalizing with $G = 1$ and $c = 1$

[89] Fred Cooperstock and Steven Tieu "General relativistic velocity: the alternative to dark matter", in: *Modern Physics Letters A* vol. 23 (2008): 1748

[90] The equation (103.1) in L. Landau & E. Lifshitz, *The Classical Theory of Fields,* 4ª revised ed. (2002):339 is the equation (9) in Fred Cooperstock and Steven Tieu, "General relativistic velocity: the alternative to dark matter", in: *Modern Physics Letters A* vol. 23 (2008): 1748

[91] Fred Cooperstock, *General Relativistic Dynamics* (2009): 142

$$e^{\lambda} = \frac{(r')^2}{1+E(R)} \qquad (59)$$

$$\text{and} \qquad \dot{r}^2 = E(R) + \frac{F(R)}{r} \qquad (60)$$

where $E(R)$ and $F(R)$ are functions of integration. This leads to the following average radial velocity equation:

$$\frac{dr}{dt} = -\frac{(\alpha+\beta)(1-\beta^2)}{8\pi r^2 \rho^2}\left[\frac{\alpha}{F} + \beta\left(\frac{F''}{(F')^2} - \frac{1}{2F}\right)\right]^{-1}\frac{\partial\rho}{\partial t} \qquad (61)$$

which "*stands in sharp contrast to the very simple Newtonian-like expression*"[92]

$$v_{local} = -\beta = -\sqrt{\frac{F}{r}} \qquad (62)$$

$$\text{and where} \quad \alpha = \frac{rF'}{3F} \qquad (63)$$

$$\text{and} \qquad \rho = \frac{F'}{8\pi r' r^2} \qquad (64)$$

The factor F is the "*accumulated mass function*"[93] conceived as a function of the radius R of the galaxy cluster (wherein the average radial velocity is supposed to be known)

$$F(R) = k_1 R^{k2} \qquad \text{and} \qquad M(R) = F(R)/2 \text{ (note 94)} \qquad (65)$$

For example, in the case of the Coma cluster, Cooperstock has the following values of the mass function: $F = 6.641*10^{-16}R^{1.453} \Rightarrow F' = 9.649*10^{-16}R^{0.453} \Rightarrow F'' = 4.371*10^{-16}R^{-0.547}$

These equations permit us to reconstruct the relation between radial velocity, galaxy cluster mass and galaxy mass density, in a relativistic model, without necessity of non-baryonic dark matter. For example, in the case of the Coma cluster, the radial velocity expressed as the ratio $2M(R_0)/r_0$ is of the order of 10^{-4} "*if we assume, as would a Newtonian, that there exists dark matter present to account for the observed velocities*" and of the order of 10^{-5} "*if we accept only the existence of the matter that we see*."[95] Now the problem we face is whether we can reconcile the 'observed velocities' and the 'baryonic matter that we see', without resorting to dark matter. Cooperstock argues that we can, with the help of the relativistic radial velocity equation (61) and the accumulative mass function of BOX 6. In stead of boosting the mass of the galaxy by adding dark matter, we boost the velocity based on visible baryonic matter using relativistic gravitational dynamics. Assuming the baryonic, visible mass is 20% or 30% or 40% of the *supposed* total mass within a sphere of 3 Mpc of $1.3*10^{15}$ solar masses, we obtain a 'boost factor' n of the *supposed* Newtonian radial velocity associated with only observed baryonic mass ($dr/dt = -n\beta$) of $n = 2.23$, $n = 1.82$ and $n = 1.58$, respectively, to obtain the relativistic

[92] Fred Cooperstock, *General Relativistic Dynamics* (2009): 146
[93] Fred Cooperstock, *GeneralRrelativistic Dynamics* (2009): 149
[94] Equations (10.26) and (10.19), respectively in Fred Cooperstock, *General Relativistic Dynamics* (2009): 149,143
[95] Fred Cooperstock, *General Relativistic Dynamics* (2009): 148

radial velocity. Since the observed average radial velocity and all the terms at the right hand side of the relativistic radial velocity equation (61) are known, we can obtain the value of the change of mass density over time ($\partial\rho/\partial t$), that is $2.13*10^{-41}\,kg/m^3/s$, $2.62*10^{-41}\,kg/m^3/s$ and $3.02*10^{-41}\,kg/m^3/s$, respectively. "*Rates of density changes of the order of magnitude* $10^{-41}\,kg/m^3/s$ *are quite reasonable as over a period of one billion years*",[96] which is the time of the evolution of the Como galaxy cluster. Cooperstock concludes that, while this is only one example, "*and a very rough one at that, ... we have been able to account for the observed velocities of galaxies within a cluster .... solely within the framework of general relativity and without any extraneous dark matter*."[97]

Cooperstock comments his findings:

"*When the gravity was deduced to be weak within these clusters, astronomers naturally turned to Newtonian gravity to correlate the seemingly anomalously large galactic velocities that they measured with the masses that they believed to be present. In this manner they initially deduced that there must be unseen "dark matter" in the order of 100 times as much as the visible matter to make the mass totals accord with the veloicities. However, with the later discovery of very large quantities of gaseous matter, this figure was reduced dramatically but there still remained a large quantity of matter yet to be accounted for. This apparent need is still promoted vigorously by researchers throughout the world. It has spawned a plethora of paspers advocating new particles that would conceivable play the role of this exotic missing material. However, we have seen that, insofar as high rotational velocities of stars in galaxies as the basis for the need for dark matter is concerned, the replacement of Newtonian gravity by general relativity removes this requirement. An essential point is that the nonlinearities of general relativity play an important role in systems of freely falling gravitating masses, leading to expressly non-Newtomian behavior, even when the gravitational field is weak. (...) Had Zwicky done this calculation 70 years ago with general relativity in mind, he might have come to very different conclusions regarding the requirement for vast stores of exotic dark matter.* "[98]

4.2.- Brownstein & Moffat

Brownstein and Moffat too presented a relativistic model of galaxy clusters explaining their radial velocity and total mass without the necessity of non-baryonic dark matter.[99] They do not start with the geodesic equation, as Cooperstock and Tieu do, but with Newton's laws of acceleration and gravitation, transformed by Einstein's general relativity. In Mathematical Box (8), I present their physical-mathematical argument and thereafter, by way of illustration of the results, I reproduce the graph of the galaxy cluster Coma, which permits us to compare the total mass estimates in the Newtonian and relativistic gravitational models.

---

[96] Fred Cooperstock, *General Relativistic Dynamics* (2009): 152
[97] Fred Cooperstock, *General Relativistic Dynamics* (2009): 152
[98] Fred Cooperstock, *General Relativistic Dynamics* (2009): 148, 153.
[99] Joel Brownstein and John Moffat, "Galaxy Cluster Masses Without Non-Baryonic Dark Matter", in: *Monthly Notices of the Royal Astronomical Society* (2005): 1-16

MATHEMATICAL BOX 7. RELATIVISTIC ESTIMATE OF THE MASS OF GALAXY CLUSTERS ACCORDING TO BROWNSTEIN AND MOFFAT

Brownstein and Moffat apply general relativity to the gravitational dynamics of 106 galaxy clusters that emit X-ray radiation that had been previously analyzed by Reiprich and Böhringer with Newtonian gravitational parameters. Brownstein and Moffat part from a Riemannian pseudo metric tensor and a third rank skewed symmetric tensorial field, called *metric-skew-tensor-gravity*. The cluster mass derived from their relativistic model is M_{MSTG}. On the other hand, the same cluster mass derived from Newtonian dynamics is M_N. The mathematical argument permits comparing both mass estimates. The Newtonian acceleration is:

$$a_N(r) = -\frac{G_0 M(r)}{r^2} \qquad (66)$$

so that the total Newtonian mass estimate is:

$$M_N(r) = -\frac{a(r)\, r^2}{G_0} \qquad (67)$$

For a spherical, isotropic and isothermal gas cloud, the acceleration, in both Newtonian and relativistic models, is:

$$a(r) = -\frac{3\beta k T}{\mu m_p}\left(\frac{r}{r^2 + r_c^{\,2}}\right) \qquad (68)$$

From equations (67) y (68), we obtain the total cluster mass equation in the Newtonian model:

$$M_N(r) = -\frac{3\beta k T}{\mu m_p G_0}\left(\frac{r^3}{r^2 + r_c^{\,2}}\right) \qquad (69)$$

and since the relativistic acceleration is:

$$a(r) = -\frac{G(r)\, M_{MSTG}(r)}{r^2} \qquad (70)$$

the total cluster mass in the relativistic model is:

$$M_{MSTG}(r) = -\frac{a(r)\, r^2}{G(r)} \qquad (71)$$

From equations (68) and (71), we obtain the total cluster mass in the relativistic model:

$$M_{MSTG}(r) = -\frac{3\beta k T}{\mu m_p G(r)}\left(\frac{r^3}{r^2 + r_c^{\,2}}\right) \qquad (72)$$

Combining the equations (69) and (72), we obtain the relationship between the total mass estimates in the Newtonian and relativistic models:

$$M_{MSTG}(r) = \frac{G_0}{G(r)} M_N(r) \qquad (73)$$

In other publications, Moffat and Brownstein obtained the equation for the gravitational constant in a very large galaxy cluster with radius r:[100]

$$G_\infty \equiv \lim_{r>>r_0} G(r) = G_0 \left\{ 1 + \sqrt{\frac{M_0}{M_{gas}}} \right\} \qquad (74)$$

From equations (73) and (74), we obtain:

$$M_{MSTG} = \left\{ 1 + \sqrt{\frac{M_0}{M_{gas}}} \right\}^{-1} M_N \qquad (75)$$

The values of r_0 and M_0 are constant:

$$r_0 = r_{out}/10 \quad \text{for } r_{out} \leq 650\,kpc \qquad (76\ A)$$

$$r_0 = 139.2\,kpc \quad \text{for } r_{out} > 650\,kpc \qquad (76\ B)$$

$$\text{and} \quad M_0 = 58.8 * 10^{14} M_{SOL} \left(\frac{M_{gas}}{10^{14} M_{SOL}} \right)^{0.39} \qquad (77)$$

From equation (75), the reader may appreciate that the relativistic total mass estimate is much less than the Newtonian estimate, which allows us to get rid of the speculations on 'missing mass' and non-baryonic dark matter', according to Brownstein and Moffat: "*we have used the simplest [relativistic] isotropic β-model based upon hydrostatic equilibrium to fit the X-ray galaxy cluster data without the need for exotic dark matter*."[101]

The following graph of Brownstein and Moffat allows us to appreciate the difference between the relativistic and Newtonian total mass estimates of the Coma galaxy cluster. The difference between the relativistic and Newtonian total mass estimates is equivalent to the Newtonian non-baryonic dark matter estimate, so that, in the relativistic model, the non-baryonic dark matter speculation is not needed. The authors reproduce similar results for another 105, X-ray radiating, galaxy clusters from the Reiprich and Böhringer sample.[102]

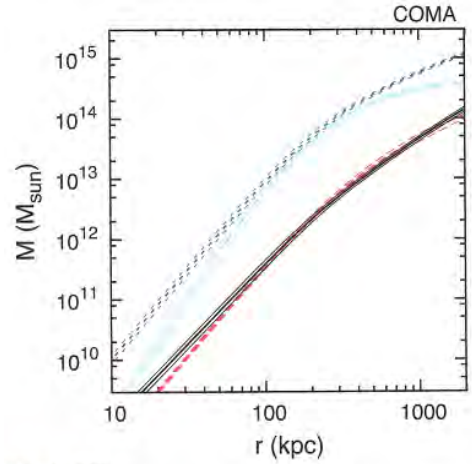
---

[100] Joel Brownstein & John Moffat, "Galaxy Rotation Curves Without Non-Baryonic Dark Matter", arXiv:astro-ph/0506370 (2005); and John Moffat, "Gravitational Theory, Galaxy Rotation Curves and Cosmology without Dark Matter", arXiv:astro-ph/0412195 (2005) and "Scalar-Tensor-Vector Gravity Theory", arXiv:gr-qc/0506021 (2005)

[101] Joel Brownstein and John Moffat, "Galaxy Cluster Masses Without Non-Baryonic Dark Matter", in: *Monthly Notices of the Royal Astronomical Society* (2005): 5

[102] Joel Brownstein y John Moffat, "Galaxy Cluster Masses Without Non-Baryonic Dark Matter", in: *Monthly Notices of the Royal Astronomical Society* (2005): 8-16

Graph.- Relativistic and Newtonian total mass estimates of the galaxy cluster COMA [103]



We saw above the case of the *Bullet Cluster 1E0657-558*, the mass of which could not be explained, within Newtonian gravitational dynamics, without the presence of non-baryonic dark matter, with a non-baryonic / baryonic mass rate of $M_{NB} / M_B = 3.17$, according to Clowe and his team.[104] However, Brownstein and Moffat have proven that in relativistic gravitational dynamics, the dark matter hypothesis is not needed. The following table gives the results for the *Bullet Cluster* in the different models.

TABLE.- NEWTONIAN AND RELATIVISTIC TOTAL MASS ESTIMATES OF THE *BULLET CLUSTER 1E0657-558*

Type of matter	**Newtonian model Clowe *et al.*** [105]	**Relativistic model Brownstein & Moffat** [106]
Baryonic	24 %	100 %
-ICM gas	No estimate given	83 %
-visible stars (galaxies)	No estimate given	17 %
Non-baryonic	76 %	0 %

[103] Joel Brownstein y John Moffat, "Galaxy Cluster Masses Without Non-Baryonic Dark Matter", in: *Monthly Notices of the Royal Astronomical Society* (2005): 7

[104] See Section 1

[105] Douglas Clowe *et al*., "A direct empirical proof of the existence of dark matter", arXiv:astro-ph/0608407, reproduced thereafter inn: *Astrophysical Journal Letters* (2006). For big galaxy clusters, Böhringer estimates that the baryonic mass, on average 15% of total mass, is distributed between the interstellar medium (13%) and stars (2%), and non-baryonic dark matter is 85% (See Section 1).

[106] Joel Brownstein & John Moffat, "The Bullet Cluster 1E0657-558 shows Modified Gravity in he Absence of Dark Matter", arXiv:astro-ph/0702146. Moffat's and Brownstein's Modified *Gravity* model (MOG) is not to be confused with Milgrom's MOND model, because the former is a relativistic model and the latter, Newtonian (See mathematical BOX 6).

Part 5.- Concluding remarks from the point of view of philosophy of science.-

It is time for a conclusion, in the words of Cooperstock: "*For the most part, astronomers continue to ignore general relativity in making deductions from observations. Thus an industry has arisen of massive computer simulations with billions of conjectured dark matter particles. The claim has been made that these simulations confirm that the CDM (cold dark matter) model of structure formation is in accord with observed structures in galaxy surveys such as the Sloan Digital Sky Survey. However, the basis for these simulations is Newtonian gravity. The lesson from our work is that the best theory of gravity, general relativity, is capable of providing surprises,*" in making the dark matter hypothesis superfluous.[107]

We can put this conclusiobn in terms of *Popper's philosophy of science*. Popper defines refutability as the demarcation between scientific theories and not-scientific theories. Let us see the following example:

a) Universal statement: "*all swans are white*".
b) Basic statement that refutes the universal statement in a specific space-time region: "*right here and now we observe this black swan*".
c) Existential statement: "*black swans exist*."

Logically, the verification of the basic statement (b) refutes the universal statement (a) and verifies the existential statement (c). However, the statement "*no black swans to be seen here*", which refutes the basic statement (b), does not refute the existential statement (c), nor does it corroborate the universal statement (a), because there may be other space-time regions where black swans do exist. This is why we say that universal statements can be refuted, but cannot be corroborated, and existencial statements can be corroborated, but cannot be refuted.

TABLE.-REFUTABILITY AND NON-REFUTABILITY OF 3 KINDS OF STATEMENTS

	UNIVERSAL STATEMENT	**BASIC STATEMENT**	**EXISTENTIAL STATEMENT**
Refutable by the facts	**Yes**	**Yes**	**No**
Verifiable by the facts	**No**	**Yes**	**Yes**

Since both universal and basic statements are refutable, they are both scientifc. However, existential, or metaphysical, or theological statements or those of science fiction cannot be refuted by the facts of the physicsal world and are therefore not scientific, which does not mean that they cannot be useful. For example, philosophy of science is metaphysics, its statements cannot be refuted by the facts, but it is very useful. ***The frontier between scientific and not-scientific statements, asccording to Kartl Popper, is the refutability principle.***[108]

I will now add a few points to Popper's philosophy of science. He did not take into account the possibility that two theories, one scientific and one speculative, can both be corroborated by the facts. What would be the demarcation principle in such a case? Let me first explain the

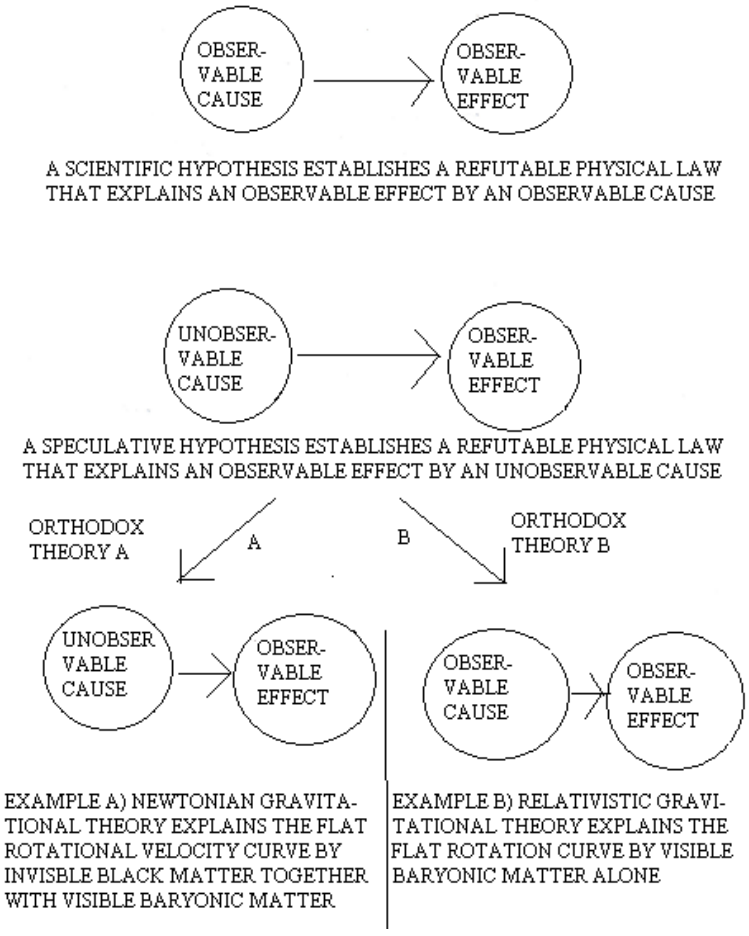
[107] Fred Cooperstock, *General Relativistic Dynamics* (2009): 159
[108] Karl Popper, *Conjeturas y Refutaciones* (1989):63-64

difference between scientific and speculative statements. A scientific hypothesis establishes a physical law that causally relates an observable cause and andIan observable effect. A speculative hypothesis establishes a physical law that explains an observable effect by an unobservable cause. For example, the hypothesis that explains galaxy rotational velocity curves by dark matter in a Newtonian gravitational theory is a speculative hypothesis. And the hypothesis that explains the flat rotational velocity curves of spiral galaxies in a relativistic gravitational theory is a scientific hypothesis.

Speculative hypotheses are necessary in the history of science. After some time, however, with the advance of scientific theories and/or scientific observations, a speculative hypothesis may end up competing with a scientific theory, where the originally unobservable cause can be substituted by an observable cause. It is my view that *when two explanations, derived from two different, but orthodox scientific theories, compete with each other, we have to give preference to the orthodox theory that postulates an observable cause over the orthodox theory that postulates an unobservable cause*.

The following scheme explains these different possible developments of a speculative hypothesis

Graph.- Scientific and speculative hypothesis



Part 6.- The origin of the speculation on dark energy

Modern cosmology supposes that, on a large scale, the Universe is flat, that is to say, $\Omega_{tot} = 1$ and $\Omega_k = 0$. Furthermore, empirical observations seem to indicate that $\Omega_M \cong 0.3$.[109] The difference between Ω_{tot} and Ω_M is usually explained by the speculation on dark energy and the dark energy density Ω_Λ. This interpretation is based on the ΛCDM model, that has been developed and refined during the last ten years,[110] and that implies the assumptions of Newtonian gravitational dynamics and a homogeneous Universe. I evaluate the dark energy hypothesis in Part 6, and in Part 76, I analyze the alternative Buchert-Wiltshire model, which reached a certain degree of maturity only two years ago.[111] The latter model makes the speculation on dark energy superfluous, and is based on the assumptions of relativistic gravitational dynamics and an inhomogeneous Universe. This new model is capable of explaining the same astrophysical observations as the ΛCDM model, that is to say the acceleration of the expansion of the Universe; the evolution of large structures over time; and the anisotropies of the Cosmic Microwave Background Radiation ($CMBR$), though from a radically different theoretical perspective,.

Part 6 has the following sections:
1) the evidence in support of the hypothesis on the acceleration of the expansion of the Universe
2) Bayesian probability calculus
3) the evolution of the gravitational dynamics of galaxy clusters
4) the form of the anisotropies of the cosmic background radiation ($CMBR$)
5) the attempts at theoretical explanation of the nature of dark energy

6.1.- Dark energy and the acceleration of the expansion of the Universe

Kirshner points out that the dark energy in Guth's speculation about early inflation is not the same dark energy as the one driving the recent acceleration of the expansion of the Universe: "*a large dollop of dark energy whose negative pressure drove the inflation era and another, much longer-lived dark energy that drives cosmic acceleration now*."[112] At the end of the 1990s evidence was presented in favor of the conjecture about the recent acceleration of the expansion of the Universe, in a local close by region (on the basis of the model with $\Omega_M \approx 0.3, \Omega_\Lambda \approx 0.7$). I am speaking of the observations of luminosity and redshift of supernovae type 1a, discovered in the *Supernova Cosmology Project* of Saul Perlmutter and his team[113] and the *High-z Supernova Search Team*, of Robert Kirshner and Adam Riess and their team.[114] According to Kirshner, the

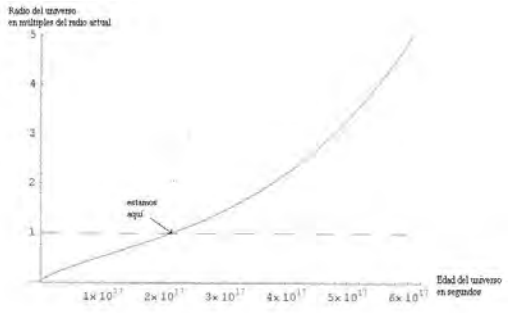
---

[109] See Section 13.5, in John Auping, *El Origen y la Evolución del Universo* (2009)
[110] Joshua Frieman, Michael Turner & Dragan Huterer, "Dark Energy and the Accelerating Universe", arXiv:astro-ph/0803.0982 (2008)
[111] David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages," in: *New Journal of Physics* (2007)
[112] Robert Kirshner, *The Extravagant Universe* (2002): 138, mis negrillas
[113] Saul Perlmutter, "Medidas de Omega y Lambda de 42 supernovas de gran corrimiento al rojo", en: *Astrophysical Journal* vol. 517 (1999): 565-586
[114] Adam Riess, "Prueba observacional de las supernovas para un Universo en aceleración y una constante cosmológica," en: *Astronomical Journal* , vol. 116 (1998): 1009-1038; y Robert Kirshner, *The Extravagant Universe* (2002)

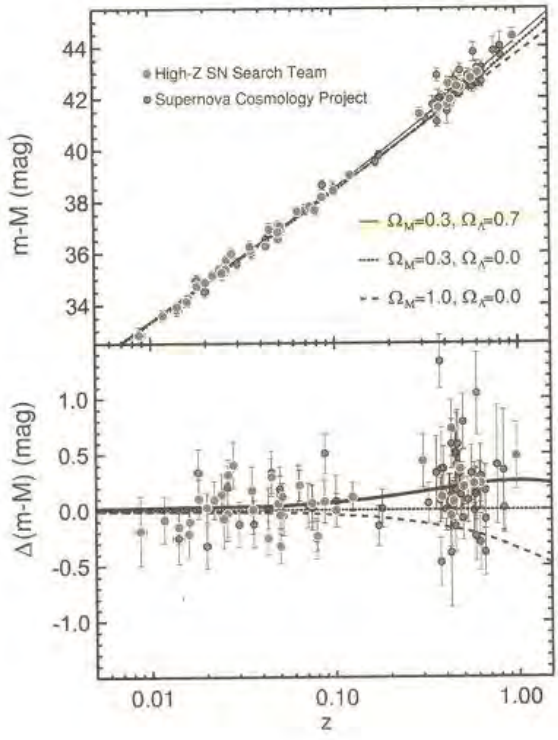
estimate of the distance of supernovae on the basis of their redshift, supposes a Hubble constant of $\pm 70/s/Mpc$, and implies a reduction of the error margin of 40 to 70%.

According to these data, type 1a supernovae that are relatively close by have a redshift that is larger than would be expected in the case of a decelerating expansion of the Universe, indicating that in the last thousands of millions of years the expansion is accelerating. Many cosmologists attribute this recent expansion acceleration to a modern edition of the ancient cosmological constant, first proposed by Einstein. [115] The sum of the mass density of $\Omega_M \approx 0.3$ and the dark energy density of $\Omega_\Lambda \approx 0.7$ yields a total density of $\Omega_{tot} = 1$, as can be seen in the next graph.

Graph.- Computer simulation of the Friedmann-Lemaître model with cosmological constant



Graph.- The apparent acceleration of the expansion of the Universe [116]



[115] Robert Kirshner, *The Extravagant Universe* (2002): 223
[116] Image reproduced in Robert Kirshner, *The Extravagant Universe* (2002): 223

The same data presented by the Kirshner-Riess and Perlmutter teams led Karttunen and his colleagues, in 2003, to a more cautious interpretation: "*a choice of model cannot be made on the basis of these observations.*"[117] In 2001, Robinson had made the same point, that is, that the original supernovae type 1a observations can be consistent with $\Lambda = 0.7$ and $\Omega_{tot} = 1$, as well as with $\Lambda = 0$ and $\Omega_{tot} = 0.3$.[118]

However, years later both teams presented again observations of the same phenomenon, but this time more precise ones, made with the *Hubble Space Telescope* and reduced the error margin of the observed luminosity considerably. In 2003, Knop and Perlmutter and their team presented data of 11 supernovae of high redshift observed by the *Hubble Space Telescope*[119] and in 2004, Riess and Kirshner and their team used the same telescope for more precise observations of 16 recent supernovae and revaluated the past evidence of 170 supernovae type 1a and affirmed having corroborated once again the hypothesis of the recent acceleration of the expansion of the Universe.[120] They also affirmed that the historical transition from deceleration to acceleration occurs at a distance that corresponds to a redshift of $z = 0.46 \pm 0.13$. In 2005, Astier and his team published their estimates of the cosmological parameters on the basis of observations of 71 high redshift type 1a supernovae discovered during the first year of the *Supernova Legacy Survey* (SNLS) that will last a total of five years.[121]

6.2.- Baysesian probability calculus.

Recently, in the cosmological literature, the concept has emerged of "model-independent cosmology" in recognition of the fact that the supposed corroboration of certain interpretations of observational data depend on certain values of the model's parameters which in themselves are dependent on the truth of certain assumptions, for example, the validity of Newtonian gravitational dynamics, and the homogeneity of the Universe. The interpretations are "model dependent". First, certain parameter values are programmed in the computer software and then the computer produces results that are compatible with these assumptions and validate the interpretations. From the point of view of logic, these model-dependent interpretations are not really *corroborated* by the data but rather shown to be *compatible* with them, and other interpretations, based on other models, might also be compatible with the same data.

In 2005, Moncy John, an astrophysicist from India, was among the first to propose "*a model-independent, cosmographic approach to cosmology,*"[122] using Bayesian probability calculus.

---

[117] Hannu Karttunen and others, *Fundamental Astronomy, Fourth Edition* (2003): 374

[118] Michael Robinson, *Los nueve números del Cosmos* (2001): 172

[119] Rob Knop, *et al.,* "New constraints on Ω_M, Ω_Λ, and w from an Independent Set of Eleven High-Redshift Supernovae Observed with the HST " (2003), arXiv:astro-ph/0309368

[120] Adam Riess *et al*., "Type Ia Supernova Discoveries at z<1 From the Hubble Space Telescope: Evidence for Past Deceleration and Constraints on Dark Energy Evolution", arXiv:astro-ph/0402512 and *Astrophysical Journal*, vol. 607 (2004): 665-738

[121] Pierre Astier *et al*., "The Supernova Legacy Survey: Measurement of Ω_M, Ω_Λ and w from the First Year Data Set", arXiv:astro-phy/0510447 (2005)

[122] Moncy John, "Cosmography, Deceleration Past, and Cosmological Models: Learning the Bayesian Way", in: *The Astrophysical Journal* vol. 630 (2005): 667

This method permits calculating the probability that certain data, based on empirical observations are generated in a Universe that corresponds to a certain theoretical model. At the same time, one calculates the probability that the same data correspond to another theoretical model of the Universe. Then one compares these different probabilities, in order to decide which model has a greater probability of not being false. This decision does not impede two important facts, in the first place, the fact that various theoretical models are compatible with the same data, and in the second place, the fact that as a result of the variation of these empirical data in different samples, different samples (for example, of type 1a supernovae), can generate different levels of probability for the same theoretical model.

Using the data of the Knop-Perlmutter and Riess-Kirshner teams, Moncy John discovered that "*the Bayesian analysis shows that ... there is no evidence from supernovae data to conclude that a changeover from deceleration to acceleration occurred anywhere in the past* $5*10^{17}$ *s* ."[123] The past $5*10^{17}$ seconds are the last 15 thousand million years, that is, the total age of the Universe. A second important conclusion of Moncy John is "*that the present analysis rules out neither the accelerating nor the decelerating models; instead we can safely conclude that the data cannot discriminate between these models*."[124]

MATHEMATICAL BOX 8, BAYESIAN PROBABILITY CALCULUS[125]

"Bayesian evidence" $E(M)$ in favor of some cosmological model M is defined as the probability P that certain empirical data D are observed in a sample, in the case that this model would be the one that corresponds to the physical reality of the Universe:

$$E(M) = P(D|M) \qquad (78)$$

and the Bayes factor is the rate of the Bayesian evidence for both models M_i y M_j:

$$B_{ij} = \frac{E(M_i)}{E(M_j)} \qquad (79)$$

If $B_{ij} > 1$, we prefer model M_i over model M_j and vice-versa, if $0 < B_{ij} < 1$, we prefer model M_j over M_i. Then we draw the natural logarithm of the Bayes Factor, in order to compare the different models M_i and M_j with a basic model M_0:

$$if\ 0 < B_{0i} = \frac{E(M_0)}{E(M_i)} < 1 \ \Rightarrow \ \ln B_{0i} < 0 \qquad (80)$$

$\Rightarrow$ so we prefer M_i over M_0;

$$if\ \ B_{0i} = \frac{E(M_0)}{E(M_i)} = 1 \ \Rightarrow \ \ln B_{0i} = 0 \qquad (81)$$

$\Rightarrow$ both models, M_i and M_0 are plausible;

[123] Moncy John, "Cosmography, Deceleration Past, and Cosmological Models: Learning the Bayesian Way", in: *The Astrophysical Journal* vol. 630 (2005): 672

[124] Moncy John, "Cosmography, Deceleration Past, and Cosmological Models: Learning the Bayesian Way", in: *The Astrophysical Journal* vol. 630 (2005): 672

[125] Harold Jeffreys, *The Theory of Probability,* 3rd edition (1998)

$$if\ B_{0i} = \frac{E(M_0)}{E(M_i)} > 1 \ \Rightarrow \ \ln B_{0i} > 0 \qquad (82)$$

$\Rightarrow$ we prefer M_0 over M_i.

Of course, $|\ln B_{0i}|$ values close to zero make the model comparison *inconclusive*, as can be appreciated in the following table from Roberto Trotta.[126] For example if the Bayes factor is $\ln B_{0i} \cong 0.139762 \Rightarrow (P(M_0) = 1.15P(M_i)$:

$\lvert\ln B_{0i}\rvert$	$B_{0i} = P(M_0)/P(M_i)$	Probability	Strength evidence
<1.0	$\leq 3:1$	<0.750	Inconclusive
1.0	$\approx 3:1$	0.750	Weak evidence
2.5	$\approx 12:1$	0.923	Moderate evidence
5.0	$\approx 150:1$	0.993	Strong evidence

This inductive method is analogous to the χ^2, where the observed distribution is compared to the expected distribution under the null hypothesis, and a decision is made whether this difference is statistically significant.[127]

Following the road initiated by Moncy John, Elgaroy and Multamäki analyzed two type 1a supernovae samples, that is, the Riess-Kirshner —also called the 'Gold' sample by Elgaroy and Multamäky— and the Astier sample (SNLS).[128] The base model M_0 with which both samples are compared is a flat Universe ($k = 0$) with a constant slightly negative deceleration factor q and linear expansion,[129] which differs from the ΛCDM model that has a transition from deceleration to acceleration. There are various surprising results of this Bayesian analysis:

1.- In the Gold sample, the more probably true model, is a closed Universe ($k = +1$), with a slight and constantly negative deceleration parameter ($q_0 = -0.04$).[130] The model that comes second is a flat Universe ($k = 0$), also constantly accelerating ($q_0 = -0.29$).[131]

2.- In the SNLS sample, the model most probably true is a flat Universe ($k = 0$), also with constant acceleration ($q_0 = -0.42$).[132] In the second place comes a model of a flat Universe

---

[126] R. Trotta, "Bayes in the sky: Bayesian inference and model selection in cosmology," en: arXiv:0803.4089, p. 14

[127] See Philip R. Bevington & D. Keith Robinson, *Data reduction and error anslysis for the physical sciences* (2003)

[128] Øystein Elgaroy & Tuomas Multamäki, "Bayesian analysis of Friedmannless cosmologies," arXiv:astro-ph/0603053

[129] $q(z) = \sum q_i z^i$ .The authors define $q = -\frac{1}{H^2}\frac{\ddot{a}}{a}$, which, given that $H = \frac{\dot{a}}{a}$, implies $q = -\frac{a\ddot{a}}{\dot{a}^2}$, so that if $q < 0 \ \Rightarrow \ \ddot{a} > 0$ (acceleration of the expansion) and if $q > 0 \ \Rightarrow \ \ddot{a} < 0$ (deceleration).

[130] $Ln(B_{0i}) < 0;\ q_0 = -0.04;\ \chi^2 = 191.1$. The authors do not explain how a closed Universe can have a slightly negative deceleration, which is a slightly positive acceleration.

[131] $Ln(B_{00}) = 0;\ q_0 = -0.29;\ \chi^2 = 182.8$

[132] $Ln(B_{00}) = 0;\ q_0 = -0.42;\ \chi^2 = 112.0$

($k = 0$), with a non-constant and non-linear deceleration parameter and a transition from deceleration q_2 to acceleration q_0 and q_1: ($q_0 = -0.60; q_1 = -0.60; q_2 = +0.61$).[133]

3.- However, in both samples the most probably true model has a constant linear expansion, and this fact implies there was never a transition from deceleration to acceleration in the expansion of the Universe, according to Elgaroy and Multamäki:

"[T]he best model in both cases has $q(z)$ constant. It therefore seems fair to conclude that there is no significant evidence in the present supernovae data for a transition from deceleration to acceleration, and claims to the contrary are most likely an artifact of the parameterization used in the fit of the data. (...) It is at the moment not possible to say anything about when, or indeed if the Universe went from deceleration to acceleration. "[134]

4.- Another important conclusion is "*that the two samples do not enable us to draw conclusions about the underlying model, [since] there is no evidence that anything beyond a constant, negative deceleration parameter is required in order to describe the data.*"[135]

Shapiro and Turner follows followed the same line of reasoning as Moncy John and Elgaroy and Multamäki.[136] They showed that the interpretation of supernovae type 1a data by the Riess-Kirshner, Knop-Perlmutter and Astier teams, start from certain assumptions that are not necessarily true. Shapiro and Turner show that other interpretations are possible, for example, "*a long epoch of recent deceleration is consistent with the data at the 10% [confidence] level"* y "*the present SNe 1a data cannot rule out the possibility that the universe has actually been decelerating for the past 3 Gyr [=three thousand million years] (i.e., since $z = 0.3$).*"[137] If we abandon the assumption, that is part of the standard ΛCDM model, [138] that the Universe is flat, another interpretation is possible, that is, "*a positively curved universe with constant negative acceleration [=a closed universe, with constant deceleration] fits the gold set surprisingly well, and allowing q [=the deceleration parameter] to vary does not significantly improve the fit*."[139]

In synthesis, the Bayesian analysis realized by Moncy John, Elgaroy and Multamäki, and Shapiro and Turner, reveals that certain empirical data are *compatible* with the standard ΛCDM model, but this compatibility does not *corroborate this model*, because other possible models

---

[133] $Ln(B_{0i}) = 0.6$; $q_0 = -0.60; q_1 = -0.60; q_2 = +0.61$; $\chi^2 = 110.5$. These data do not imply that the M_i model (flat Universe, with a transition from deceleration to acceleration) is the more probably true one, but, on the contrary, the M_0 model with constant negative deceleration is almost twice as probably true as the M_i model.

[134] Øystein Elgaroy & Tuomas Multamäki, "Bayesian analysis of Friedmannless cosmologies," arXiv:astro-ph/0603053, p.5, 6

[135] Øystein Elgaroy & Tuomas Multamäki, "Bayesian analysis of Friedmannless cosmologies," arXiv:astro-ph/0603053, p.7

[136] Charles Shapiro & Michael Turner, "What do we really know about cosmic acceleration?", in: *Astrophysical Journal* vol. 649 (2006): 563-569

[137] Charles Shapiro & Michael Turner, "What do we really know about cosmic acceleration?", in: *Astrophysical Journal* vol. 649 (2006): 566, mis negrillas

[138] In this model, the parameter for the equation of state has a value of -1 ($w = P/\rho_{VAC} = -1$).

[139] Charles Shapiro & Michael Turner, "What do we really know about cosmic acceleration?", in: *Astrophysical Journal* vol. 649 (2006): 568

exist, which, with different degrees of probability of not being false, can explain the same observational data.

6.3.- Dark energy and the gravitational dynamics of galaxy clusters

In 2006, Longair pointed out that the observation of large scale structures, as, for example, galaxy clusters in different stages of the evolution of the Universe, do no permit distinguishing between different cosmological models, at the present moment of the history of the Universe. On the basis of some super computer simulations, run by Guinevere Kauffmann and her team,[140] he compared the evolution of large scale structures in four types of universes, among them: [141]

1.- The same large scale structures generated in a flat Universe with cosmological constant, in the ΛCDM standard model, so that $\Omega_0 + \Omega_\Lambda = 1$, with $\Omega_\Lambda = 0.7$.

2.- Large scale structures generated in an open Universe without a cosmological constant, $OCDM$ (=*open cold dark matter*), with an overall density parameter of approximately $\Omega_0 = 0.3$.

3.- τCDM Large scale structures generated in an open Universe without a cosmological constant with cold dark matter and decaying neutrinos.

According to Longair, there is no difference in the results of the dynamics of the subjacent model, if we compare the ΛCDM model (with $\Omega_\Lambda = 0.7$) with the $OCDM$ and τCDM models (with $\Omega_\Lambda = 0$). He argues that *"this is because the dynamics only differ from* $\Omega_\Lambda = 0$ *(...) in the late stages of evolution of the Universe when the effect of the cosmological constant is to stretch out the timescale of the model, allowing some further development of the perturbations."*[142] The structures with a larger redshift are farther away from us in space and time. For that reason, if we go from the right to the left in the following images, from $z = 0$ to $z = 3$, the size of the same object is diminishing progressively,[143] and the angle from which we observe the object is progressively smaller. The important fact is that the ΛCDM *, on the one hand,* and the $OCDM$ and τCDM series of images, on the other hand, the first one with and the second and third ones without the cosmological constant, *are identical*. There do not yet appear differences due to the cosmological constant, in none of the four stages of the evolution of the Universe that are contemplated. According to Longair, the differences would appear in later stages of its evolution that we have not yet reached.

The original authors of these images, Guinevere Kauffmann and her team, say exactly the same as Longair comparing ΛCDM ($\Omega_M = 0.3; \Omega_\Lambda = 0.7$) and τCDM ($\Omega = 1$): "*Although neither model is perfect both come close to reproducing most of the data. Given the uncertainties in*

---

[140] Guinevere Kauffmann *et al*., "Clusters of galaxies in a hierarchical Universe, in: *Monthly Notices of the Royal Astronomical Society,* vol. 303 (1999): 188-206

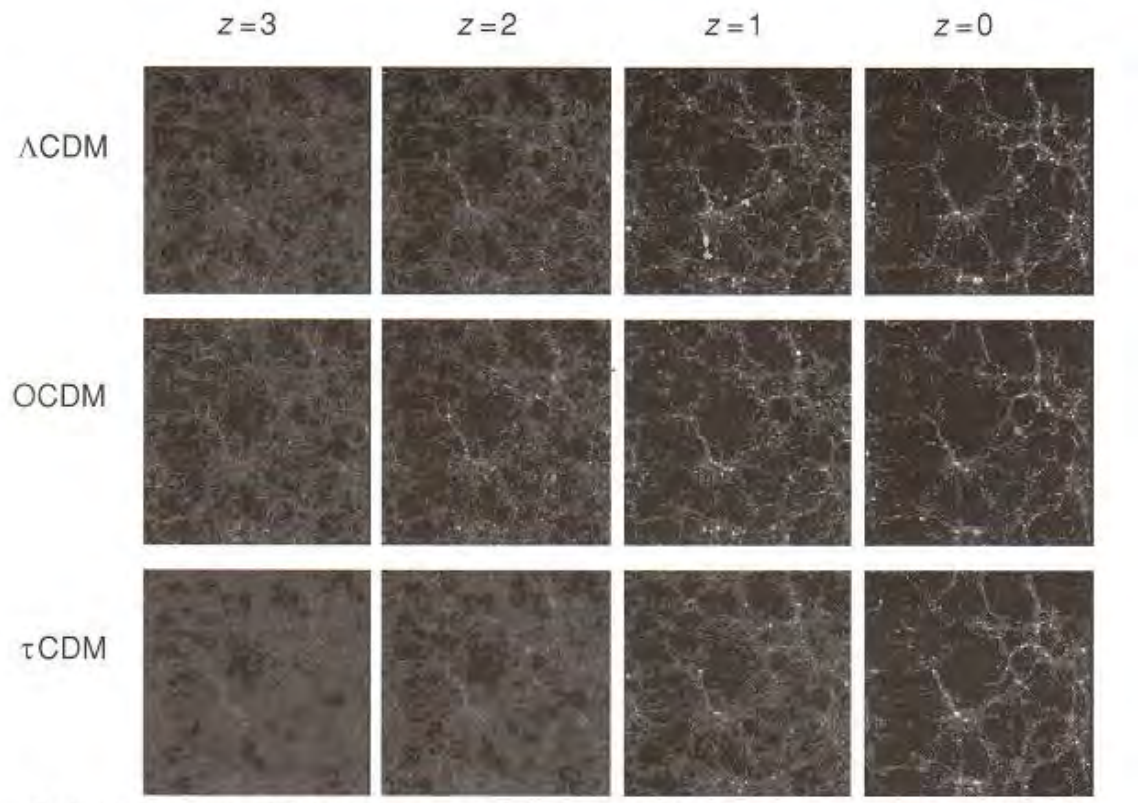
[141] Malcolm Longair, *The Cosmic Century* (2006): 414

[142] Malcolm Longair, *The Cosmic Century* (2006): 415, my underlining.

[143] If we take the age of the Universe as a function of the matter density and the redshift of light, we can calculate the different values of z. See Edward Kolb & Michael Turner, *The Early Universe* (1994): 504, whose equations yield $t = 2.0571 * 10^{17} (\Omega_0 h^2)^{-1/2} (1+z)^{-3/2} s$, ($\Omega_0 = 0.28; h = 0.7$) $\rightarrow t = 17.6$ thousand million years for $z = 0$; $t = 11.1$ thousand million years for $z = 1$; $t = 8.5$ thousand million years for $z = 2$; and $t = 7.0$ thousand million years for $z = 3$.

modeling some of the critical physical processes, we conclude that it is not yet possible to draw firm conclusions about the values of cosmological parameters from studies of this kind." [144]

Graph.- Computer models of the Universe with and without cosmological constant [145]



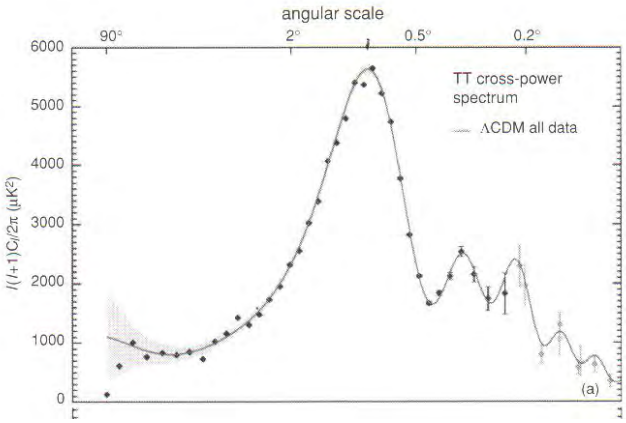
Explanation: firts series: LCDM (flat Universe with dark energy and cold dark matter; second series; open universe with cold dark matter, but no dark energy; third series: open universe with cold dark matter and neutrino decay but no dark energy

[144] Guinevere Kauffmann, Jörg Colberg, Antonaldo Diafero & Simon White, "Clustering of Galaxies in a Hierarchical Universe: I. Methods and Results at z=0", in: *Monthly Notices of the Royal Astronomical Society*, vol. 303 (1999): 188-206 (quote on p. 288; also arXiv:astro-ph/9805283, 21 May 1998) & "II. Evolution to High Redshift", ibidem, vol. 307 (1999): 529-536 (also arXiv:astro-ph/9809168, 18 September 1998). My underlining.
[145] Malcolm Longair, *The Cosmic Century* (2006): 414

6.4- Dark energy and the cosmic microwave background radiation *CMBR*

Modern cosmology analyzes the *CMBR* from different angles, in order to measure the small variations or anisotropies of that radiation, that were first discovered by George Smoot and his team.[146] Supposedly, these anisotropies prove that the Universe has a flat geometry, which would imply that dark energy exists. By way of example, I reproduce a graph to be found in Longair. The curve is generated by the computer, with he help of a Legendre function, while the software has been programmed with the values of the parameters of some predetermined cosmological model. In the case of the graph reproduced below, the assumption is made that the ΛCDM model, with non zero cosmological constant, is the valid one, with a matter density of $\Omega_M \approx 0.3$ and a dark energy density of $\Omega_\Lambda \approx 0.7$. The black points in the graph represent the observations: in the horizontal axis we read the angle from which the observations are made, and in the vertical axis one reads the magnitude of the observed anisotropies. The maximum variation in the temperature is one in hundred thousand.

Graph.- The ΛCDM model is compatible with the anisotropies of the cosmic background radiation *CMBR* [147]



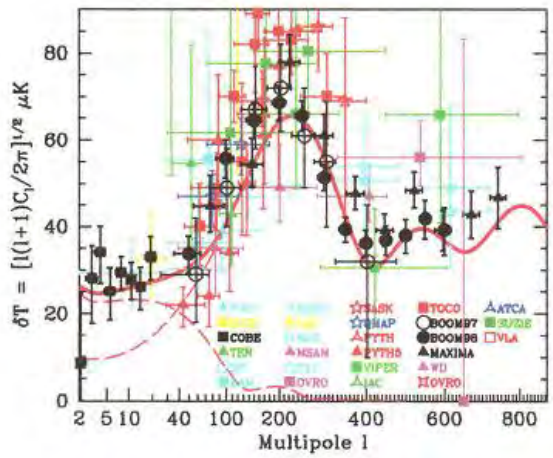
If one looks at this graph, one may think "well, there is no doubt here, the facts corroborate the theory". Things are not as simple, however, as they appear to be at first sight. In the first place (A), these observations (the dots) are a kind of 'average' of many observations that yield wildly varying results among themselves. In the second place (B), the curve (the continuous line) is model-dependent.

A.- Uncertainties.- The following graph of Tegmark, Zaldarriaga and Hamilton, published in the year 2000, presents the observations made by 27 different teams of cosmologists. The curve that best fits these discrepant observations is the solid red line, a kind of 'average' of the 27 different series of data. It represents a cosmological model with a closed Universe ($\Omega_{tot} = 1.3$),[148] which is of course quite different from the standard ΛCDM, where the Universe is flat ($\Omega_{tot} = 1.0$). For this

[146] George Smoot & Keay Davidson, *Wrinkles in Time* (1993)
[147] Malcolm Longair, *The Cosmic Century* (2006): 424
[148] Max Tegmark, Matías Zaldarriaga y Andrew Hamilton, "Towards a refined concordance model: joint 11-parameter constraints from CMB and large scale structure", arXiv:astro-ph/0008167 (2000): p. 1

reason, according to the authors, there is a need for complementing these observations of the *CMBR* with other observations, to be able to arrive at the standard ΛCDM model.

Graph.- The discrepancy between 27 different observations of the anisotropies in the *CMBR*



B.- Model dependency.- In this graph of the year 2000, the cosmological model that lies at the basis of the interpretation of the observational data is defined by 11 parameters. However, Uros Seljak and Matías Zaldarriaga of Harvard perfected the model, and their 2009 version has 16 cosmological parameters, assumed to be true, among them the densities of cold dark matter, and dark energy, and another 20 non-cosmological parameters, which makes a total of 36 assumptions. In the table below, I reproduce the 16 cosmological parameters of the *CMBFast* model of Seljak and Zaldarriaga.[149]

TABLE.- 16 COSMOLOGICAL PARAMETERS IN THE CMBFAST MODEL

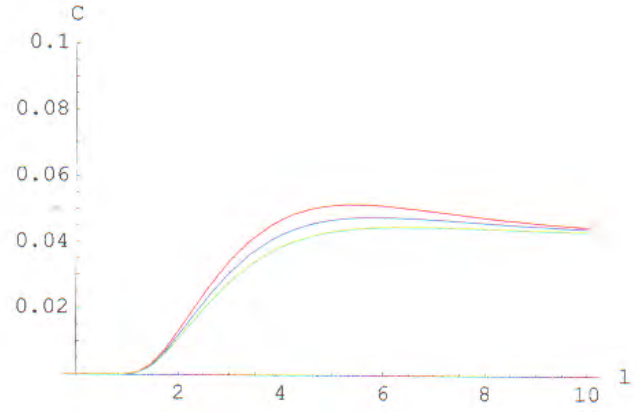
Cosmological parameters

wdyn	Equation of state for the dark energy component
Omega_b, Omega_c, Omega_de,Omega_nu	baryons, CDM, Dark Energy and massive neutrino contributions to the energy density
H0	Hubble constant in km/sec/Mpc
Tcmb	Present CMB temperature
Y_He	Helium abundance
N_nu(massless) (= *annur*)	Number of massless neutrinos
N_nu (massive) (= *annunr*)	Number of Massive neutrinos
g*(massive)	g* (number of relativistic degrees of freedom) when massive neutrinos decouple
n (= *an*)	scalar spectral index
alpha_n=dn/dlnk (= *alphans*)	running of the scalar spectral index
nt (= *ant*)	tensor spectral index
alpha_nt=dn/dlnk	running of the tensor spectral index
rat	tensor to scalar ratio
optdlss	Optical depth to the last scattering surface
zri	redshift of reionization
rfi	ionization fraction after reionization

[149] Uros Seljak & Matías Zaldarriaga, "List of CMBFAST parameters," CMBFAST Website

If we vary, by way of example, just one of these 16 cosmological parameters, that is, the assumption of homogeneity (the blue line), the maximum anisotropy varies considerably (red and green lines) when we assume a non-homogeneous Universe. [150]

Graph.- Variation of $\pm 20\%$ in the observed maximum anisotropy (red and green line) when the assumed homogeneity of the Universe (blue curve) is abandoned



Explanation: Edward Kolb & Michael Turner, *The Early Universe* (1990): 384, 386

$$\left\langle |a_{lm}|^2 \right\rangle = \frac{AH_0^{n+3}}{16} \frac{\Gamma(3-n)}{\Gamma[(4-n)/2]^2} \frac{\Gamma[(2l+n-1)/2]}{\Gamma[(2l+5-n)/2]} \qquad (83)$$

$$\left\langle \frac{\delta T(\hat{x}_1)}{T} \frac{\delta T(\hat{x}_2)}{T} \right\rangle = \frac{1}{4\pi} \sum_{l=2}^{\infty} (2l+1) \left\langle |a_{lm}|^2 \right\rangle P_l(\hat{x}_1 . \hat{x}_2) \exp\left[-(l+1/2)^2 \sigma^2\right] \qquad (84)$$

Let me explain the graph. The letter "**l**" in the horizontal axis represents the angle of observation (the order of the polynomial of Legendre) and "**C**" in the vertical axis represents the correlation between intensities and the anisotropy of the *CMBR*. The variation of the maximum peak of the curve is due to the variation of the constant "**n**" that represents the degree of homogeneity, the blue line representing perfect homogeneity. This is just an example of how the selection of the values of the parameters affects the curve and, for that reason, can be used to fit the observational data.

In other words, the assumptions of the model used to interpret the observational data, make these data *compatible* with the concordant cosmological model but do *not corroborate* it. Analogously, we may have 20 different economic models to explain a certain rate of inflation. They may all succeed in doing so, though they may be wildly different among themselves. Of course, if we demand that a model explain not one or two but, simultaneously two thousand different types of observational data, at many points in time, the restraints on the model become more severe. If such a model would succeed in doing so, we may say that the data corroborate the theoretical model.

This model dependency, added to the considerable discrepancy of the observations by different teams, makes it impossible to draw definite conclusions about the underlying cosmological

[150] Computer simulation run by Alfredo Sandoval and John Auping with Wolfram's *Mathematica*, using the equations 9.144 and 9.148 of Edward Kolb & Michael Turner, *The Early Universe* (1990): 384, 386

model. The data do not refute one model and corroborate another, but only prove that different models, among them the standard ΛCDM are *compatible* wit the data. Given this model dependency and the error margins in each of the 27 observations that vary considerably,[151] it comes as no surprise that different models may succeed, with different degrees of probability, in different samples, to predict the same observational data, as I explained in the foregoing section on Bayesian probability.[152]

From the point of view of the philosophy of science, we can formulate the same problem in yet another way. When the observational data fit a theoretical model with 36 parameters, which of these are really compatible with the data? We might increase the value of one parameter and decrease the value of another one and the same data may fit the 'new' model. Of course, not all parameters are that free. Some are not free at all, because they were corroborated independently as part of some other theory. For example, the value of the primordial and the present helium abundances —one of the 36 parameters of the model—, are known with a considerable degree of certainty.[153] However, the values of other parameters are completely model-dependent. Among the latter ones we may count the assumed homogeneity of the Universe, the temperature T of the $CMBR$, the value of the Hubble constant H_0, the matter density Ω_M and the space curvature Ω_k. As we shall see below, in relativistic models some of these parameters may obtain very different values and others are completely dropped, among them the supposed dark energy density Ω_Λ.[154]

[151] See above, the graph "The discrepancy between 27 different observations of the anisotropies in the $CMBR$"
[152] See part 6.2
[153] See Section 13.5 of John Auping, *El Origen y la Evolución del Universo* (2009)
[154] See Part 7

Part 7.- General relativity refutes the speculation about dark energy

In some textbooks, the expansion of the Universe is compared with an inflating globe, while coins stuck on its surface move away from each other. The expansion of the globe is conceived to be homogeneous and symmetric, so that the same rate of expansion is observed in all directions. Most cosmologists too use the assumption of the homogeneity and isotropy of the Universe to construct their models of its expansion. This assumption does not appear to be true. Even though the Cosmic Microwave Background Radiation (*CMBR*) reveals that the Universe, some 300,000 years after the *Big Bang*, was almost perfectly homogeneous and isotropic, it is a fact that the small inhomogeneities present at that time —as registered by the small anisotropies of that radiation—, have since been magnified on a very large scale and *today* the Universe is not homogeneous, but rather an ensemble of enormous voids surrounded by enormous walls of galaxy clusters, like an enormous sponge.

Peebles analyzed the problem of the small scale inhomogeneity of the Universe, but concluded that there is still evidence in favor of the assumption of large scale homogeneity and isotropy of the Universe,[155] though he admitted certain biases do occur due to irregular mass distributions:

"*The clumpy mass distribution in the real world can play quite different roles in different tests. (...) Irregular mass distribution can produce a systematic error in apparent magnitudes of galaxies, for the mass along the line of sight acts as a lens that determines the rate of change of the convergence of a bundle of light rays, and that fixes the angular size of the image. If the mass distribution is clumpy rather than smooth (...) observations could be biased to favor objects whose images have been magnified, because they appear brighter, or the bias could go the other way, for where there is mass there tends to be dust. Thus, if the line of sight to a distant object passes through a large amount of mass, so that gravitational lensing is magnifying the angular size of the object, the image tends to be obscured.*" [156]

Peebles thought these biases would cancel each other out, so that the overall "*bias may not be large.*"[157] This may be true for the kind of inhomogeneities Peebles was contemplating, but there appear to be others, not considered by him, that produce important overall systematic biases, especially the so called *backreaction*, and the different rates at which clocks are running in voids and walls, which do affect the value of the parameters of the model, as we shall see below.

Our galaxy cluster is located in an enormous void of 200 to 300 Mpc that expands between 20 and 30% more rapidly than could be expected according to the global (average) Hubble constant[158]; there is a superstructure of 400 Mpc known as *Sloan*'s *Great Wall*, surrounding part of this void; more locally there are two other minor voids of 35 to 70 Mpc each and *Shapely*'s super cluster with a diameter of 40 Mpc, at a distance of some 200 Mpc from our galaxy; and

---

[155] James Peebles, "Fractal Universe and Large-Scale Departures from Homogeneity" and "Cosmology in an Inhomogeneous Universe", in: *Principles of Physical Cosmology* (1993): 209-224, 343-360

[156] James Peebles, *Principles of Physical Cosmology* (1993): 343

[157] *Ibidem,* p. 343

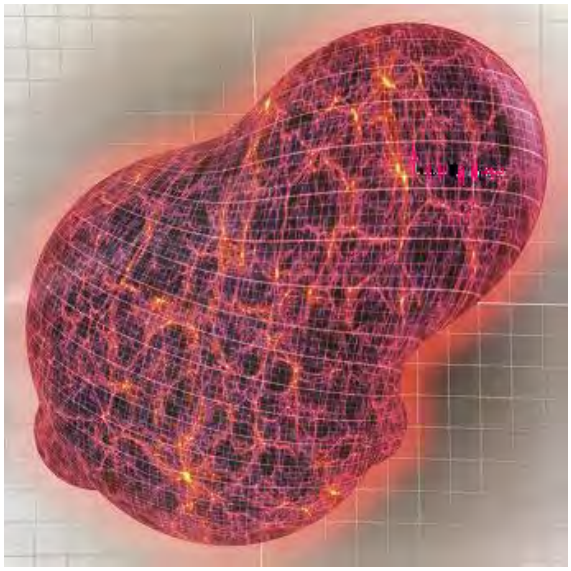
[158] Paul Hunt & Subir Sarkar, "Constraints on large scale inhomogeneities from WMAP-5 and SDSS: confrontation with recent observations", arXiv:0807.4508, pág.1

according to the *Hubble Space Telescope Key Project* there exists a significant anisotropy in the local expansion, at distances of up to 100 Mpc ([note 159]).

In general, Wiltshire estimates that "*some 40-50% of the volume of the universe at the present epoch is in voids of* $30h^{-1}$ *Mpc [*$\approx$*between 40 a 50 Mpc] in diameter (...) and there is much evidence for voids 3 to 5 times this size, as well as local voids on smaller scales*."[160] Obviously, with the passing of time, due to the expansion of the Universe, the participation of the voids in the total volume increases, and that of the gravitationally collapsed regions decreases.

Today, the expansion of the Universe looks more like a river with rapids than a slowly inflating globe. Instead of a homogeneous flow of a slow river, the flow of the rapids makes the total flow inhomogeneous. Some parts flow more rapidly than others and between slow and faster flows friction occurs. Where there are obstacles the current surrounding them is slowed down and vortices are generated in the surface of the water. The expansion of the Universe resembles such a current that meets obstacles in the form of galaxy clusters and black holes that slow down the rate of expansion of some regions, and the friction between faster and slower expanding regions generates shear, while enormous vortices are generated by black holes. Walls of galaxy clusters with the size of hundreds of Megaparsecs surround voids where the expansion is more rapid, than in regions with higher matter density. The Universe is more like a expanding and deforming sponge than a smoothly and homogeneously expanding globe, something resembling the image produced in a recent article in the *Scientific American*.[161]

Image.- The deceleration of the expansion of the Universe is greater in the walls than in the voids



[159] Nan Li & Dominic Schwarz, "Scale dependence of cosmological backreaction", arXiv:astro-ph/0710.5073, p. 1
[160] David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in *New Journal of Physics* (2007): 4
[161] Timothy Clifton & Pedro Ferreira affirm in "Dark Energy: Does it really exist?", in: *Scientific American*, vol. 300 (2009): 33, that we are in the centre of an enormous spherical void with the size of the observable Universe, diameter $850 \pm 170 h^{-1}\ Mpc$. See also their "Living in a void", arXiv:0807.1443 (2008): 1-4.

7.1.- How averaging parameters in a non-homogeneous Universe produces the *backreaction*

The Universe is a collection of regions with high matter density (*walls*), where local clocks run slowly, and regions with low matter density (*voids*), where clocks run faster. We can disappear this inhomogeneity through a process of averaging also known as *smoothing*, but that does not take away the fact that the values of cosmological parameters in *walls* and *voids* differ among themselves and from the global-average. This real inhomogeneity invalidates one of the basic assumptions of the standard model and affects the curvature of space. Even though the Universe has a global-average curvature, that does not take away the differences between gravitationally collapsed regions, like galaxies and galaxy clusters, where space does not expand and the curvature is positive ($k > 0$), especially close to the galaxy centers where black holes can be found, and the enormous voids in the Universe, where space expands faster, and the curvature is negative ($k < 0$). The curvature of space is not homogeneous.

A numerical example helps to understand the process of smoothing. Let us choose 27 random numbers between 01 and 100,[162] with a maximum difference between the biggest and the smallest number of 94. We then average every three numbers and obtain nine new numbers[163] that differ less among themselves than the original 27, with a maximum difference of 59 between the extreme values. We now repeat the process of averaging groups of three numbers, and obtain three new numbers, [164] with a maximum difference among themselves of 33.0. If we average these three numbers we obtain the global-average of 48.6.[165] That last number is analogous to the global-average of the cosmological parameters and the initial differences between the 27 original numbers is analogous to the differences between the local values of the cosmological parameters. The process of getting from the global inhomogeneity to the final, global average is the process of smoothing. In that process, the inhomogeneity of the Universe totally disappears.

How can we obtain the global-average values of the cosmological parameters? Obviously, that could be done in theory by obtaining a weighed average of their values in different regions of the Universe, of its voids and walls. There is a complication, however, since the values of its parameters evolve and change with time, so that their magnitude is not constant, neither on the local scale, nor on the large scale. Consequently, we have two options, the first one of which would be to obtain the average of the original values of the parameter in different regions at the beginning of the Universe, and then see how this average evolves. The second option would be to let the parameter evolve with time in different regions of the Universe and obtain an average of these independent evolving values in the final stages of the evolution of the Universe. Normally, the operation of averaging and the operation of evolving in time are commutative, so that the same result is obtained, independently of the order in which these two operations are executed, as can be appreciated in the following mathematical box.

---

[162] The random numbers are taken from Hubert Blalock, *Social Statistics* (1960): 437, the numbers are: 10, 09, 73, 25, 33, 76, 52, 01, 35, 86, 34, 67, 35, 48, 76, 80, 95, 90, 91, 17, 39, 29, 27, 49, 45, 37 y 54.
[163] The averages are: 30.7, 44.7, 29.3, 62.3, 53, 88.3, 49, 35 y 45.3.
[164] The averages are: 34.9, 67.9 y 43.1.
[165] The mean is: 48.6.

MATHEMATICAL BOX 9. THE OPERATIONS OF AVERAGING AND DERIVING OVER TIME ARE COMMUTATIVE

Normally, the operations of obtaining the average or the derivative are commutative. In the first equation, we first obtain the average and then the derivative over time:

$$\partial_t \left\langle x^3 + 2x^2 - 3x + 6 \right\rangle = \partial_t (\frac{1}{4}x^3 + \frac{1}{2}x^2 - \frac{3}{4}x + 1.5) = \frac{3}{4}x^2 + x - \frac{3}{4} \quad (85)$$

and in the second one, we first obtain the derivative and then the average:

$$\left\langle \partial_t x^3 + \partial_t 2x^2 - \partial_t 3x + \partial_t 6 \right\rangle = \left\langle 3x^2 + 4x - 3 + 0 \right\rangle = \frac{3}{4}x^2 + x - \frac{3}{4} \quad (86)$$

but even so, in both cases, we obtain exactly the same result. This is what we mean when we say that even in the case of non-linear equations the operations of obtaining the average or the derivative are commutative.

The problem with many cosmological parameters is that they are determined by tensorial equations, where averaging and deriving over time are NOT commutative operations. It is not the same to let an average matter distribution and its corresponding spatial geometry evolve in time, or let the matter distributions of different regions and their corresponding spatial geometries evolve in time and then average the final results. Cosmologists tend to first average matter distributions and its corresponding geometries and then use Einstein's equations to obtain the homogeneous geometry that results from the evolution in time of this average. Actually, the proper procedure would be to first resolve Einstein's equations for the different geometries of the different regions of the Universe, then let these results evolve in time, and then average the final results. Since these operations are not commutative, not following the proper order of operations yields erroneous results, according to Wiltshire, referring to previous work of Buchert: "*the geometry which arises from the time evolution of an initial average of the matter distribution does not generally coincide, at a later time, with the average geometry of the full inhomogeneous matter distribution evolved via Einstein's equations.*"[166]

The first one to draw attention to the fact that the operations of averaging and resolving Einstein's equations are not commutative, was George Ellis, in 1984.[167] He showed that the structure of the non-linear equations of general relativity is substantially modified by the process of large scale *smoothing*. Let us see this point first graphically (see next graph) and then algebraically (see mathematical BOX 10). The following graph taken from Ellis represents three scales in measuring phenomena in the Universe, that is, the scale of stars and solar systems, the scale of galaxies; and the scale of galaxy clusters and walls of galaxy clusters.[168]

Obviously, the evolution of the Universe in time has taken place in the opposite order, starting with a homogeneous cloud of hydrogen and helium, which existed some 300,000 years after the

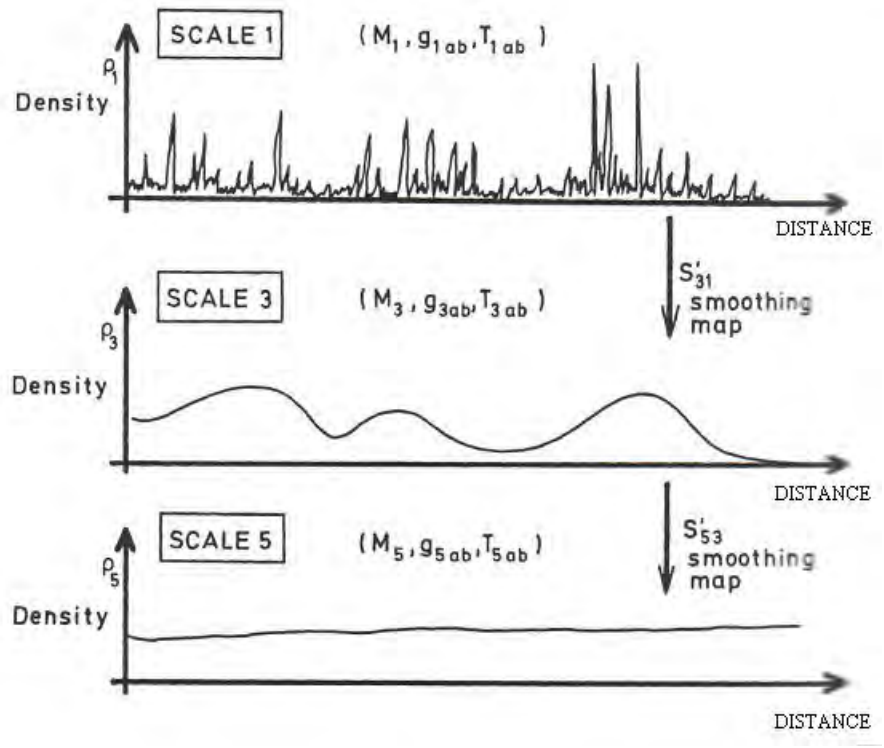
[166] David Wiltshire, "Exact solution to the averaging problem in cosmology", arXiv:0709.0732

[167] George Ellis, "Relativistic Cosmology: Its Nature, Aims and Problems", in: B. Bertotti *et al*., eds., *General Relativity and Gravitation*, págs. 215-288.

[168] George Ellis, "Relativistic Cosmology: Its Nature, Aims and Problems", in: B. Bertotti, *et al.,* eds., *General Relativity and Gravitation* (1984): 230

Big Bang, up to the voids and walls of galaxies and galaxy clusters, and the stars and solar systems that we observe today, in our inhomogeneous Universe.

Graph.- The operation of averaging cosmic phenomena on an ever bigger scale: scale 1 = details down to stars; scale 3 = galaxies; scale 5 = large scale features



Ellis showed that the operation of averaging and resolving Einstein's tensorial equations are not commutative: "*Thus, a significant problem at the foundation of cosmology is to provide suitable definitions of averaged manifolds*[169] *(..), of metric* $[G_{\mu\nu}]$ *and stress-tensor* $[T_{\mu\nu}]$ *averaging and smoothing procedures, and to show these have appropriate properties*"[170]

We get from the local scale to the global scale by averaging or smoothing. The problem is, as we can see below, in mathematical BOX 10, that in order to get from scale 1 to scale 3, and from scale 3 to scale 5, the tensorial equation that is valid on scale 1, is no longer valid on scale 3, or scale 5. In order to correct the error that occurs when we first average the matter density and its corresponding geometry in different regions of the Universe, and then solve Einstein's equation, we will have to introduce a term of correction, also known as the *backreaction*, in the tensorial equations used on scale 3 and on scale 5.

[169] "Manifolds" are multiples of different space-time regions of the Universe

[170] George Ellis, "Relativistic Cosmology: Its Nature, Aims and Problems", in: B. Bertotti, *et al.*, eds., *General Relativity and Gravitation*, p. 231.

MATHEMATICAL BOX 10 THE BACKREACTION TERM IN EINSTEIN'S TENSORIAL EQUATION

Einstein's tensor is:[171]

$$G^{\mu\nu} = R^{\mu\nu} - \frac{1}{2} g^{\mu\nu} R = \kappa T^{\mu\nu} = -\frac{8\pi G}{c^4} T^{\mu\nu} \quad (87)$$

and on scale 1, this tensor look as follows:

$$G_{1\mu\nu} = R_{1\mu\nu} - \frac{1}{2} g_{1\mu\nu} R_1 = \kappa T_{1\mu\nu} \quad (88)$$

but on scale 3 and 5, the left hand term of the equation, representing the average of the metric, no longer is equal to the right hand term, representing the mass-energy average:

$$G_{3\mu\nu} \neq \kappa T_{3\mu\nu} \quad (89)$$

$$G_{5\mu\nu} \neq \kappa T_{5\mu\nu} \quad (90)$$

We therefore need, in equations (89) and (90) a term also known as the *backreaction,* $P_{3\mu\nu}$ y $P_{5\mu\nu}$, which leaves Einstein's tensor as follows:[172]

$$G_{3\mu\nu} = R_{3\mu\nu} - \frac{1}{2} g_{3\mu\nu} R_3 = \kappa T_{3\mu\nu} + P_{3\mu\nu} \quad (91)$$

$$G_{5\mu\nu} = R_{5\mu\nu} - \frac{1}{2} g_{5\mu\nu} R_5 = \kappa T_{5\mu\nu} + P_{5\mu\nu} \quad (92)$$

The theoretical terms of the *backreaction*, $P_{3\mu\nu}$ y $P_{5\mu\nu}$, are analogous to the term Q_D of Buchert, Kolb, Matarrese and Riotto in BOX 11, 15 and 16, but Ellis did not define its empirical value.[173] Zalaletdinov, an astrophysicist of Uzbekistan, has given us a precise mathematical definition of Einstein's 'average' tensorial equation, improving, in his view, the previous work by Buchert.[174] Einstein's tensor, applicable on all scales, according to Zalaletdinov, is the following:[175]

$$\bar{g}^{\beta\varepsilon} M_{\beta\gamma} - \frac{1}{2} \delta^{\varepsilon}_{\gamma} \bar{g}^{\mu\nu} M_{\mu\nu} = -\kappa \left\langle T^{\varepsilon(micro)}_{\gamma} \right\rangle + (Z^{\varepsilon}_{\mu\nu\gamma} + \frac{1}{2} \delta^{\varepsilon}_{\gamma} Q_{\mu\nu}) \bar{g}^{\mu\nu} \quad (93)$$

where $\bar{g}^{\beta\varepsilon} M_{\beta\gamma}$ is the average curvature tensor and $\bar{g}^{\mu\nu} M_{\mu\nu} = M$, the average curvature scalar.

---

[171] See equation (286) of Appendix VI B of John Aupíng, *Origen y Evolución del Universo* (2009)

[172] George Ellis, "Relativistic Cosmology: Its Nature, Aims and Problems", in: B. Bertotti *et al.,* eds., *General Relativity and Gravitation*, págs. 233

[173] William Stoeger, Amina Helmi & Diego Torres, in "Averaging Einstein's Equations: The Linearized Case", arXiv:gr-qc/9904020, have made an attempt at averaging Einstein's non-linear equations in linear form.

[174] Roustam Zalaletdinov has found the exact way to average Einstein's non-linear equations in non-linear form, in many publications, of which I mention only two: "Averaging out the Einstein's Equations", in: *General Relativity and Gravitation,* vol. 24 (1992): 1015-1031; and "Averaging problem in general relativity, macroscopic gravity and using Einstein's equations in cosmology", in: *Bulletin of the Astronomical Society of India* (1997): 401-416.

[175] Roustam Zalaletdinov, "Averaging out the Einstein's Equations", in: *General Relativity and Gravitation,* vol. 24 (1992): 1025 equation (23)

Regrettably, during fifteen years, Ellis' warnings were not taken into account by cosmologists in the construction of their models. There was a general tendency to estimate the values of the global cosmological parameters at the present time, and then project them back to the origins of the Universe, in simplified models, where Newtonian gravitational dynamics were assumed to be valid at non-relativistic velocities, and the Universe was assumed to be homogeneous, from beginning to end. The assumptions and simplifications of these models did not alarm too many people, and often these assumptions were note even consciously made. However, the problems became more acute at the end of the 90's, when the apparent acceleration of the expansion of the Universe was discovered by Kirshner, Perlmutter and Riess. Only in the case that the local expansion rates were equal to the global-average expansion rate, as would be the case in a homogeneous and isotropic Universe, the magnitude of this *backreaction* term would be zero,[176] but, as we shall see below, this assumption proves to be invalid.

Thomas Buchert, a German astrophysicist working in France, followed up on Ellis's suggestions. I will first define some terms of Buchert's model, and then present his comparison of homogeneous and inhomogeneous models, both Newtonian and relativistic.[177]

1) D, a specific spatial-temporal dominion of the Universe;
2) H_D, the Hubble constant in this dominion;
3) $\langle R \rangle_D$, the average curvature of the Universe, represented by Ricci's scalar;
4) Θ_D, the expansion of the volume of this dominion (of the expansion of elements of the fluid);
5) σ, the shear or distortion of elements of the fluid by interaction with surrounding matter;
6) $d_t \langle \Theta \rangle_D$ or $\langle \Theta \rangle_D{}^\circ$, the evolution in time of the initial average of the expanding volumes of local dominions (first the initial average is calculated, then this average evolves in time);
7) $\langle d_t \Theta \rangle_D$, the final average of the expanding volumes of local dominions after they have evolved in time (first different local dominions evolve in time and then an average is obtained;
8) Q_D, 'the source' of non-linear results, also known as the '*backreaction term'* that measures the discrepancy between perfect homogeneity and the effect of existing inhomogeneities;[178]
9) a_D, the rate of expansion of a dominion of the Universe.

In the case of expanding, spherical, inhomogeneous volumes Θ, the operations of averaging and evolving in time are NOT commutative, as we saw above, and as a result, the *backreaction* Q_D is generated, that represents the difference between the average of the quantities that evolved separately in time, and the final result of the evolution of the average of the original quantities (see mathematical BOX 11).[179]

---

[176] Because, in that case, $d_t \langle \Theta \rangle_D = \langle d_t \Theta \rangle_D$, so that $Q_D = 0$. See Mathematical BOX 11

[177] The model of Newtonian (virial) gravitation is not necessary homogeneous. It is possible to construct Newtonian, inhomogeneous models, see Thomas Buchert, *On Average Properties of Inhomogeneous Cosmologies*, arXiv:gr-qc/00010556 (2000): 1-9

[178] Thomas Buchert, *On Average Properties of Inhomogeneous Cosmologies*, arXiv:gr-qc/00010556 (2000): 3

[179] Thomas Buchert, *On Average Properties of Inhomogeneous Cosmologies*, arXiv:gr-qc/00010556 (2000): 4; Edward Kolb, Sabino Matarrese & Antonio Riotto, "On cosmic acceleration without dark energy", in: *New Journal of Physics* (2006): 6; and *idem*, "On Cosmic Acceleration from Backreaction," on-line (2009): 13.

MATHEMATICAL BOX 11 HOW THE *BACKREACTION* IS DERIVED

The difference between $d_t\langle\Theta\rangle_D$ y $\langle d_t\Theta\rangle_D$ produces the quantity also known as the *backreaction* Q_D:

$$d_t\langle\Theta\rangle_D - \langle d_t\Theta\rangle_D = \left\langle(\Theta - \langle\Theta\rangle_D)^2\right\rangle_D = \left\langle\Theta^2 - 2\Theta\langle\Theta\rangle_D + \langle\Theta\rangle_D{}^2\right\rangle_D =$$
$$= \left\langle\Theta^2\right\rangle_D - 2\langle\Theta\rangle_D\langle\Theta\rangle_D + \left\langle\langle\Theta\rangle_D\right\rangle_D{}^2 = \left\langle\Theta^2\right\rangle_D - 2\langle\Theta\rangle_D{}^2 + \langle\Theta\rangle_D{}^2 = \left\langle\Theta^2\right\rangle_D - \langle\Theta\rangle_D{}^2 \quad (94)$$

$$Q_D = \frac{2}{3}\left(\frac{d}{dt}\langle\Theta\rangle_D - \left\langle\frac{d}{dt}\Theta\right\rangle_D\right) - 2\left\langle\sigma^2\right\rangle_D \quad (95)$$

Combining equations (94) y (95), we obtain:

$$Q_D = \frac{2}{3}\left(\left\langle\Theta^2\right\rangle_D - \langle\Theta\rangle_D{}^2\right) - 2\left\langle\sigma^2\right\rangle_D \quad (96)$$

The next mathematical box presents a synthesis of Buchert's model that differs from the Newtonian model only by integrating the term of the *backreaction*.

MATHEMATICAL BOX 12 SOME FRIEDMANN EQUATIONS IN BOTH HOMOGENEOUS AND INHOMOGENEOUS, RELATIVISTIC MODELS

I present some Friedmann equations, in both homogeneous and inhomogeneous, relativistic models developed by Buchert:

RELATIVISTIC, HOMOGENEOUS

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{1}{3}(8\pi G\rho) \quad (97)$$

$$\rho = \frac{3H^2}{8\pi G} \quad (99)$$

$$\frac{\ddot{a}}{a} = -\frac{4}{3}\pi G\rho \quad (100)$$

$$\partial_t\rho + 3\frac{\dot{a}}{a}\rho = 0 \quad (102)$$

RELATIVISTIC, INHOMOGENEOUS

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8}{3}\pi G\langle\rho\rangle - \frac{1}{6}\langle R\rangle - \frac{1}{6}Q_D =$$
$$= \left(\frac{\dot{a}}{a}\right)^2 = \frac{8}{3}\pi G\rho_{eff}\ (^{\text{note 180}}) \quad (98)$$

$$\frac{\ddot{a}}{a} = -\frac{4}{3}\pi G\langle\rho\rangle + \frac{1}{3}Q_D \quad (101)$$

$$\partial_t\langle\rho\rangle + 3\frac{\dot{a}}{a}\langle\rho\rangle = 0 \quad (103)$$

The solution to equation (103) is:

$$\langle\rho\rangle = \rho_0 a_0 / a^3 \quad (104)$$

[180] See Sabino Matarrese, Rocky Kolb & Toni Riotto, "On Cosmic Acceleration from Backreaction," on-line (2009) This means that, $\rho_{eff} = \langle\rho\rangle - \frac{Q_D}{16\pi G} - \frac{\langle R\rangle}{16\pi G}$ and $P_{eff} = -\frac{Q_D}{16\pi G} + \frac{\langle R\rangle}{16\pi G}$, where the terms with $\langle R\rangle$ indicate the average curvature and the terms with Q_D, the cinematic *backreaction*.

as can be seen from combining equations (103) y (104):

$$\partial_t(\rho_0\bar{a}_0\bar{a}^{-3}) + 3\dot{\bar{a}}\bar{a}^{-1}(\rho_0\bar{a}_0\bar{a}^{-3}) = 0$$

$$\Rightarrow \quad \rho_0\bar{a}_0(-3\bar{a}^{-4}\dot{\bar{a}}) + \rho_0\bar{a}_0(3\bar{a}^{-4}\dot{\bar{a}}) = 0 \qquad (105)$$

The equation of state of the *backreaction*, or 'integrability condition', only exists in the relativistic model and not in the Newtonian model. Buchert proposes:

$$\partial_t Q_D + 6\frac{\dot{a}_D}{a_D}Q_D + \partial_t\langle R\rangle_D + 2\frac{\dot{a}_D}{a_D}\langle R\rangle_D = 0 \quad (^{\text{note 181}}) \qquad (106)$$

and Kolb, Matarrese y Riotto, and Wiltshire propose (which is exactly the same):

$$\partial_t({a_D}^6 Q_D) + {a_D}^4\partial_t({a_D}^2\langle R\rangle_D) = 0\,(^{\text{nota 182}}) \qquad (107)$$

The following mathematical box compares two inhomogeneous models one Newtonian and one relativistic, according to Buchert. [183]

MATHEMATICAL BOX 13, PARAMETERS IN TWO INHOMOGENEOUS MODELS, ONE NEWTONIAN AND ONE RELATIVISTIC

We represent the Hubble constant by $H = \dot{a}/a$ and the spatial curvature $\langle R\rangle$ by the average Ricci scalar. The density parameters, according to Buchert, are:

NEWTONIAN, INHOMOGENEOUS	RELATIVISTIC, INHOMOGENEOUS
$\Omega_M + \Omega_\Lambda + \Omega_k + \Omega_Q = 1 \quad (108)$ $\Omega_M^D = \frac{8\pi G\langle\rho\rangle_D}{3{H_D}^2} \quad (110)$	$\Omega_M + \Omega_k + \Omega_Q = 1 \quad (109$ $\Omega_M^D = \frac{8\pi G\langle\rho\rangle}{3{H_D}^2} \quad (111)$ and $\langle\rho\rangle = \rho_0\bar{a}_0/\bar{a}^3 \quad (112)$ gives $\Omega_M^D = \frac{8\pi G\rho_0\bar{a}_0}{3{H_D}^2\bar{a}^3} \quad (113)$
$\Omega_\Lambda^D = \frac{\Lambda c^2}{3{H_0}^2}(^{\text{note 184}}) \quad (114)$	

[181] Thomas Buchert, "On Average Properties of Inhomogeneous Cosmologies", arXiv:gr-qc/00010556, p. 12, eq. 45

[182] See Edward Kolb, Sabino Matarrese & Antonio Riotto, "On cosmic acceleration without dark energy" in: *New Journal of Physics,* vol. 8 (2006): 7, eq. (25) & "On Cosmic Acceleration from Backreaction," online, p.16 & David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in: *New Journal of Physics*, vol. 9 (2007): 5

[183] Thomas Buchert, "On Average Properties of Inhomogeneous Cosmologies", arXiv:gr-qc/0001056 (2000): 4, 12. See also a synthesis of Buchert's model in David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in: *New Journal of Physics*, vol. 9 (2007): 9

[184] Some authors write $\Omega_\Lambda = \Lambda c^2/3{H_0}^2$ and Thomas Buchert, in "On Average Properties of Inhomogeneous Cosmologies", arXiv:gr-qc/0001056 (2000):4, writes $\Omega_\Lambda = \lambda/3{H_0}^2$. The two versions are compatible if one takes into account that some authors normalize the equations with $c = 1$. The same applies to equation (117).

$\Omega_Q^D = \frac{1}{3a_D^2 H_D^2} \int_{t1}^{t} Q_D \frac{da_D^2}{dt_1} dt_1$ (115)	$\Omega_Q^D = -\frac{Q_D}{6H_D{}^2}$ (116)
$\Omega_k^D = -\frac{k_D c^2}{H_D^2 a_D^2}$ (117)	$\Omega_k^D = -\frac{\langle R \rangle_D}{6H_D{}^2}$ (118)

Buchert showed that in the Newtonian, inhomogeneous model, the sum of the first three of the four terms found in the definition-equation $\Omega_M + \Omega_\Lambda + \Omega_k + \Omega_Q = 1$, is 0.99 (note 185) so the fourth term must be very small, that is $\Omega_Q = 0.01$ (nota 186): "*This term, which brought the higher voltage of having mastered a generic inhomogeneous Newtonian cosmology, shows no global relevance and it seems that we are drawn back to the previous state of low visibility of the standard cosmological models*".[187]

Roberto Sussman run a series of simulations with the relativistic dust model of Lemaître-Tolman-Bondi (LTB), in which he demonstrated that Buchert's backreaction term can have positive values, in regions with hyperbolic curvature, as well in elliptic dominions —either in isolation or surrounded by a hyperbolic exterior—, that suffer gravitational collapse, and are capable of producing an acceleration of the expansion of a universe without the need for dark energy.[188] He stresses, however, that we are dealing with a *qualitative* evaluation of the model, and that it is necessary, "*as a complement to this work, (...) to test numerically how large the effective acceleration, that we have shown here to exist, can be.*"[189] Besides, according to Wiltshire, "*approaches based on the exact LTB models or the exact Szekeres models*", though "*immensely useful, both as exact models for isolated systems in an expanding universe or as toy models* (...) *could only be applied to the universe as a whole if one abandoned the Copernican Principle*", which is more than Wiltshire is willing to do.[190]

Some cosmologists, notably Edward Kolb, Sabino Matarrese and Antonio Riotto felt that in an inhomogeneous, relativistic model, the effects of the inhomogeneities appear to be sufficiently large to let the terms Q_D and Ω_Q^D may be to replace Λ and Ω_Λ^D . In 2006, they elaborated Buchert's model, attributing the apparent acceleration of the expansion of the Universe to the backreactions of its gravitational perturbations, making the dark energy hypothesis superfluous:[191] "*Another possibility [different from the standard ΛCDM model] is that the Universe is matter-dominated and described by general relativity, and the departure of the expansion rate of the Einstein-De Sitter model is the result of backreactions of cosmological*

[185] Thomas Buchert, *On Average Properties of Inhomogeneous Cosmologies*, arXiv:gr-qc/00010556 (2000): 1-9
[186] However, even being so small, it has a strong influence on the evolution of the cosmological parameters in time. See Thomas Buchert, Martin Kerscher & Christian Sicka, "Backreaction of inhomogeneities on the expansion: the evolution of cosmological parameters", arXiv:astro-ph/9912347, p. 17.
[187] Thomas Buchert, *On Average Properties of Inhomogeneous Cosmologies*, arXiv:gr-qc/00010556 (2000): 13
[188] Roberto Sussman, "Conditions for back reaction and 'effective' acceleration in Lemaître-Tolan-Bondi dust models", arXiv:0807.1145 (2009)
[189] *Ibidem*, p. 33
[190] David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in: *New Journal of Physics* (2007): 6
[191] Edward Kolb, Sabino Matarrese & Antonio Riotto, "On cosmic acceleration without dark energy" in: *New Journal of Physics* vol. 8 (2008): 322

perturbations. This explanation is the most conservative, since it assumes neither a cosmological constant nor a modification of general relativity."[192]

Like Sussman, the authors admit that they have not yet been able to measure the quantitative influence of the backreactions generated by these gravitational perturbations: "*The actual quantitative evaluation of their effect on the expansion rate of the Universe would, however, require a truly non-perturbative approach, which is clearly beyond the aim of this paper.*"[193]

One year later, in 2007, David Wiltshire, an astrophysicist from New Zealand, commenting on this essay of Kolb, Matarrese and Riotto, pointed out exactly that: "*While perturbative approaches have naturally led to realization of the significance of backreaction, to account for '74% [of the matter-energy density of the universe by] dark energy', the effect of backreaction on the background of the universe would [have to] be so great that a viable quantitative model is beyond the domain of applicability of perturbation theory.*"[194] The probable magnitude of the *backreaction* is not large enough to explain the recent, apparent expansion of the Universe and, as a consequence, the backreaction does not serve as a possible substitute of dark energy.

Aseem Paranjape, an astrophysicist from India, applied the mathematical structure developed by Roustam Zalaletdinov, an astrophysicist from Uzbekistan capable of averaging Einstein's tensorial equations,[195] to the problem of the expansion of the Universe[196] and discussed his findings with Buchert and Wiltshire. He reached the same conclusion as Wiltshire in his criticism of Kolb, Matarrese and Riotto, that is to say, that the effects of the backreaction are real, but insufficient to explain the recent acceleration of the expansion of the Universe, which is normally attributed to the negative pressure of dark energy:

"• *Although technically possible, in the real world backreaction does not significantly affect the expansion history of the universe*.

• *Cosmological perturbation theory is stable against backreaction effects, well into the nonlinear regime.*

• *Dark energy cannot therefore be an effect of the backreaction of inhomogeneities.*"[197]

---

[192] *Ibidem*, p. 2

[193] *Ibidem*, p. 15

[194] David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in: *New Journal of Physics* (2007): 5-6. See also the same article, p. 22

[195] Roustam Zalaletdinov has found the exact way to average Einstein's non-linear equations in non-linear form, in many publications, of which I mention only two: "Averaging out the Einstein's Equations", in: *General Relativity and Gravitation,* vol. 24 (1992): 1015-1031; and "Averaging problem in general relativity, macroscopic gravity and using Einstein's equations in cosmology", in: *Bulletin of the Astronomical Society of India* (1997): 401-416

[196] Aseem Paranjape, "A Covariant Road to Spatial Averaging in Cosmology: Scalar Corrections to the Cosmological Equations", arXiv:0705.2380 (2007) and his thesis, *The Averaging Problem in Cosmology* (2009)

[197] Aseem Paranjape, *The Averaging Problem in Cosmology* (2009): 6

7.2.- Clocks run at different rates in voids and walls

In 2005, David Wiltshire presented an alternative theory,[198] which he called *fractal bubble model = FBM*). The name is not that important. Actually, this fractal structure of the Universe does not exist at small scales, as Peebles pointed out in 1993.[199] I prefer to think of the Universe as having the structure of a large, expanding sponge: large voids surrounded by large walls of filaments of galaxy clusters. We may drop the theory of the fractal structure of the Universe, without loosing the essence of Wiltshire's contribution, which we may call the theory of the differential running of watches in voids and walls or, briefly, 'the time-scape model'. Wiltshire pointed out that after the moment of recombination, some 300,000 years after the *Big Bang*, the imaginary watches, located in different regions of the Universe, started to differ increasingly, because in regions with high matter density, gravity makes watches run slower, and in voids, faster. Wiltshire revived one of the implications of general relativity, already explained by Einstein himself, who said: "*[L]et us examine the rate of a unit clock, which is arranged to be at rest in a static gravitational field. (..) [T]he clock goes more slowly if set up in the neighborhood of ponderable masses..*"[200] This lentification of clocks by gravity has three consequences:

1) The wavelength of light coming from objects with high mass density will be redshifted;
2) The velocity of objects that move away from observers located in a ponderable gravitational field, and the redshift of those objects' light will be higher when measured by the observer's wall clock than by its own co-moving clock or clocks in voids.
3) There is a third consequence not mentioned by Einstein, but not less important. Light from a supernova thast passes through the large void that surrounds us is more redshifted than the global-average, because the void expands faster than ghe global-average

This theory of the watches might explain the apparent acceleration of the Universe in the large void surrounding our galaxy cluster, as compared with the global-average deceleration. When the velocity of a supernova at the other end of the void is measured with our watch, or with a global-average watch, it might seem to move away from us at a faster rate then when it is measured with a watch mounted on the supernova itself, because our watch moves slower than its one.

However, a problem persisted with the solution offered by Wiltshire in 2005, as he himself observed. If we compare, with Bayesian probability, the ability of both models, that is the standard flat ΛCDM , which includes dark energy, and Wiltshire's FBM , without dark energy, to explain the same supernovae type 1a data, the ΛCDM model is more probably true than the FBM in the range of $0.2 \leq \Omega_M \leq 0.5$, which is the empirical range of our Universe.[201] The FBM was in need of serious revision.

---

[198] David Wiltshire, "Viable inhomogeneous model universe without dark energy from primordial inflation", arXiv:gr-cq/0503099 (2005)

[199] See James Peebles, "Fractal Universe and Large-Scale Departures from Homogeneity" in: *Principles of Physical Cosmology* (1993): 209-224

[200] Albert Einstein, "The Foundation of the General Theory of Relativity", in: *Annalen der Physik* vol. 49 (1916), traducido al inglés en *The Collected Works of Albert Einstein,* vol. 6 (1989): 197-198 (my underlining)

[201] Benedict Carter, Ben Leith, Cindy Ng, Alex Nielsen & David Wiltshire *et al*., "Type IA supernovae tests of fractal bubble universe with no cosmic acceleration", arXiv:astro-ph/0504192.

7.3.- The new relativistic Buchert-Wiltshire paradigm

The problem just mentioned was not resolved until 2007, when Wiltshire proposed his *time-scape model* that integrated Buchert's *backreaction* and his own theory on differential clock rates in an inhomogeneous universe.[202] This new, integrated model is capable of explaining the apparent acceleration of the expansion of the Universes and other phenomena that have motivated many cosmologists to accept the speculative concept of dark energy to explain them. Wiltshire distinguishes three times or imaginary clocks, that is, slow clocks in gravitationally dense and collapsed regions that measure τ_w (w for walls), rapid clocks in the voids with time τ_v (v for voids) and a global-average clock with time t ([note 203]). These three times yield three differential clock rates, that is $dt/d\tau_w$, $dt/d\tau_v$ and $d\tau_v/d\tau_w$. Only at the beginning of the Universe, at the moment of recombination, the Universe was an almost perfectly smooth and homogeneous cloud of hydrogen and helium and, as a consequence, at that time, $dt/d\tau_w \cong dt/d\tau_v \cong d\tau_v/d\tau_w \cong 1$. In order to define the passed and present-day parameters as a function of the global-average time t, Buchert's formalism is used to average the values of parameters measured with clocks in walls, τ_w and voids, τ_v. In the next mathematical box, I synthesize the Buchert-Wiltshire paradigm.

MATHEMATICAL BOX 14, THE *BACKREACTION* IN THE NEW BUCHERT-WILTSHIRE PARADIGM

The Hubble constant is defined as the ratio of the expansion velocity v_{ex} and the distance r to a particular object of the Universe ($H = v_{ex}/r$). With Wiltshire, I define the Hubble constant as a function of global-average time t, both for gravitationally collapsed regions $H_w(t)$, for the large voids $H_v(t)$ and for the entire Universe $\overline{H}(t)$. The last one is also called the subjacent or bare Hubble constant and represents a ponderated average of the former two. Obviously, the three constants have different values ($H_w < \overline{H} < H_v$).

There are two ways to obtain the bare constant, taking into account that $H_w/a_w = H_v/a_v$, where a is distance. The first equation is:

$$\overline{H}(t) = \frac{\dot{a}_w}{a_w} = \frac{1}{a_w}\frac{da_w}{d\tau_w} = \frac{1}{a_w}\frac{dt}{d\tau_w}\frac{da_w}{dt} = \frac{dt}{d\tau_w}H_w = \overline{\gamma}_w H_w =$$

$$= \frac{\dot{a}_v}{a_v} = \frac{1}{a_v}\frac{da_v}{d\tau_v} = \frac{1}{a_v}\frac{dt}{d\tau_v}\frac{da_v}{dt} = \frac{dt}{d\tau_v}H_v = \overline{\gamma}_v H_v \qquad (119)$$

where $H_w \equiv \frac{1}{a_w}\frac{da_w}{dt}$ and (26 B) $H_v \equiv \frac{1}{a_v}\frac{da_v}{dt}$ (120)

[202] *Ibidem*, p. 22, equation (32)
[203] David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in: *New Journal of Physics* (2007)

From equation (119) we obtain:

$$h_r(t) = H_w / H_v = \gamma_v / \gamma_w = d\tau_w / d\tau_v = 1/(d\tau_v / d\tau_w) \quad (121)$$

where the $d\tau_v / d\tau_w$ function is called the lapse function or the differential clock rate in walls and voids. The second way to obtain the bare Hubble constant is by way of a ponderated average of H_w and H_v:

$$H(t) = \frac{1}{3}\langle\Theta\rangle_H = f_w H_w + f_v H_v \text{ (nota 204)} \quad (122)$$

Where the factors f_w and f_v indicate the volumes of walls and voids as a proportion of the total volume of the Universe, respectively, so that::

$$f_w(t) + f_v(t) = 1 \quad (123)$$

Since, with the passing of time, the volume of the walls does not increase, but the volume of the voids increases because of the expansion of the Universe, these two factors $f_w(t)$ and $f_v(t)$, are not constant in time. Given the Ellis-Buchert formalism, the terms $\langle\Theta\rangle_H{}^2$ and $\langle\Theta^2\rangle_H$ have different magnitudes:

$$\langle\Theta\rangle_H{}^2 = 9 f_w{}^2 H_w{}^2 + 9 f_v{}^2 H_v{}^2 + 18 f_w f_v H_w H_v \quad (124)$$

$$(31)\ \langle\Theta^2\rangle_H = 9 f_w H_w{}^2 + 9 f_v H_v{}^2 \quad (125)$$

Let us remind now the backreaction term of Buchert, Kolb, Matarrese and Riotto and let us suppose, for the time being, a zero value for the shear σ. The *backreaction* is defined by the difference between $d_t\langle\Theta\rangle_D$ and $\langle d_t\Theta\rangle_D$, that is between $\langle\Theta\rangle_H{}^2$ and $\langle\Theta^2\rangle_H$, so that:

$$Q_D = \frac{2}{3}(d_t\langle\Theta\rangle_D - \langle d_t\Theta\rangle_D) - 2\langle\sigma^2\rangle_D = \frac{2}{3}\langle\Theta^2\rangle_D - \frac{2}{3}\langle\Theta\rangle_D{}^2 - 0 \quad (126)$$

By combining equations (124), (125) and (126), we obtain:

$$(33)\ Q = 6 f_w H_w{}^2 + 6 f_v H_v{}^2 - 6 f_w{}^2 H_w{}^2 - 6 f_v{}^2 H_v{}^2 - 12 f_w f_v H_w H_v \quad (127)$$

and since, by (123),

$$f_w = 1 - f_v \quad (128)$$

it follows that equation (127) can be transformed in (129):

$$Q = 6H_w{}^2 - 6 f_v H_w{}^2 + 6 f_v H_v{}^2 - 6H_w{}^2 + 12 f_v H_w{}^2 - 6 f_v{}^2 H_w{}^2 - 6 f_v{}^2 H_v{}^2 \quad (129)$$
$$-12 f_v H_w H_v + 12 f_v{}^2 H_w H_v \Rightarrow$$

$$Q = 6 f_v H_w{}^2 + 6 f_v H_v{}^2 - 6 f_v{}^2 H_w{}^2 - 6 f_v{}^2 H_v{}^2 - 12 f_v H_w H_v + 12 f_v{}^2 H_w H_v \quad (130)$$

$$\Rightarrow (37)\, Q = 6 f_v (1 - f_v)(H_v - H_w)^2 \text{ (nota 205)} \quad (131)$$

---

[204] Peter Smale & David Witshire, "Supernova tests of the timescape cosmology," en: arXiv:1009.5855v1, p.4

[205] This is the same result as obtained by David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in *New Journal of Physics* (2007): 21, equation (31), first part of the equation

We now rewrite equation (119), which is convenient because of the use we are making of it when we derive the global deceleration of the expansion of the Universe:

$$H(t) = \frac{1}{a_w}\frac{dt}{d\tau_w}\frac{da_w}{dt} \qquad (119)$$

$$\Rightarrow \frac{dt}{d\tau_w} = H(t)/\frac{da_w}{a_w\,dt} = \frac{H(t)\,a_w dt}{da_w} \qquad (132)$$

From equations (122) and (132), we obtain:

$$\frac{dt}{d\tau_w} = \frac{(f_w H_w + f_v H_v)\,a_w dt}{da_w} = \frac{f_w H_w a_w dt}{da_w} + \frac{f_v H_v a_w dt}{da_w} \qquad (133)$$

from equations (119) and (133) we obtain:

$$\frac{dt}{d\tau_w} = f_w + f_v\frac{H_v}{H_w} \quad (134)$$

and from equatios (120) and (134):

$$\frac{dt}{d\tau_w} = 1 - f_v + f_v\frac{H_v}{H_w} = 1 + f_v\left(\frac{H_v}{H_w} - 1\right) = \left(\frac{H_w/H_v + f_v(1 - H_w/H_v)}{H_w/H_v}\right) \quad (135)$$

and from (119) and (133):

$$H_w = \frac{H(t)(H_w/H_v)}{H_w/H_v + f_v(1 - H_w/H_v)} \qquad (136)$$

Since:

$$H_v = H_w/(H_w/H_v) \qquad (137)$$

it follows, from (136) and (137) that:

$$H_v = \frac{H(t)}{H_w/H_v + f_v(1 - H_w/H_v)} \qquad (138)$$

Combining (131), (136) and (138), we obtain the *backreaction* as a function of the Hubble constants that take into account the differential clock rates:

$$Q = 6 f_v(1 - f_v)\left(\frac{H^2(1 - H_w/H_v)^2}{[H_w/H_v + f_v(1 - H_w/H_v)]^2}\right) \text{ (note 206)} \qquad (139)$$

The interesting thing about equation (139) of mathematical box 14, is that there are two moments in the history of the Universe where the backreaction is zero, and the deceleration has a value of $\bar{q} = 0.0635$. At the moment of recombination, the inhomogeneities were almost zero ($h_r = 1$), because the Universe was basically a homogeneous cloud of hydrogen and helium, so that the Hubble constants of regions with different degrees of mass density had an almost identical value, generating a zero backreaction.[207] And vice-versa, when we approach the end of

[206] This is the same result as obtained by David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in *New Journal of Physics* (2007): 21, equation (31), second part of the equation

[207] $H_w/H_v = 1 \Rightarrow \Omega_Q = 0$

the history of the Universe, the volume of the voids will be so much larger than the volume with galaxy clusters, that, once more, $f_v = 1$, and once more, the backreaction will be zero.[208] At this particular moment of the history of the Universe, the deceleration is real, with $\bar{q} = +0.015348$ and the apparent acceleration of the expansion velocity measured with clocks in walls, like our galaxy cluster, is $q(\tau_w) = -0.042785$.

The new Buchert-Wiltshire paradigm has important implications for the interpretation of empirical phenomena that have defied cosmology during the last decades. Below we will see how the Buchert-Wiltshire paradigm allows us to disregard dark energy in explaining the apparent acceleration of the expansion of the Universe, the evolution of its large scale structures, and the anisotropies of the Cosmic Microwave Background Radiation *CMBR*.

The variation of the Hubble constant.- One of the most paradoxical implications of the new paradigm is the variation of the Hubble constant in walls, voids and global-average regions, and if measured with wall-clocks, void-clocks or global-average clocks.

At the beginning of 2007, Buchert pointed out that it was difficult to quantify the effects of the *backreaction.*[209] However, that same year, Nan Li and Dominic Schwarz offered some approximate estimates.[210] Below, I reproduce five important results of Li and Schwarz's study, that take into account observations made by the *Hubble Space Telescope* in the *Key Project*.

1) The effects of the *backreaction* on the variation of the Hubble constant are *scale-dependent*, that is, its variation depends on the inverse square distance ($Q_D \propto 1/r^2$).
2) At a scale of less than 200 Mpc, the influence of these inhomogeneities is much bigger in a relativistic model than in a Newtonian model.
3) The values of the variation of the Hubble constant in the relativistic model coincide with the observations of the *Key Project Hubble Space Telescope:* "*We see that the theoretical band matches the experimental data well, without any fit parameter in the panel.*"[211]
4) The Hubble constant has a comparatively larger value in our neighborhood, at a scale of about 100 Mpc, which constitutes a large void.
5) "*[C]osmological averaging (backreaction) gives rise to observable effects up to scales of* $\approx$ *200 Mpc. However, it is not sufficient to explain the observed accelerated expansion at this point.*"[212]

Other investigations go much further. Hunt and Sarkar observed that we are located in a huge void with a diameter of 200 to 300 Mpc that expands 20 to 30% faster than would be expected according to the global-average Hubble constant. The expansion acceleration observed by Hunt and Sarkar in this infra dense void is extremely improbable in the context of the standard

---

[208] $f_v \cong 1$ y $1 - f_v \approx 0 \Rightarrow \Omega_Q = 0$. See David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", en *New Journal of Physics* (2007): 28

[209] Thomas Buchert, "*Backreaction Issues in Relativistic Cosmology and the Dark Energy Debate*", arXiv:gr-qc/06112166 (enero 2007).15

[210] Nan Li & Dominic Schwarz, "Scale dependence of cosmological backreaction", arXiv:astro-ph/0710.5073

[211] Nan Li & Dominic Schwarz, "Scale dependence of cosmological backreaction", arXiv:astro-ph/0710.5073, p. 5

[212] Nan Li & Dominic Schwarz, "Scale dependence of cosmological backreaction", arXiv:astro-ph/0710.5073, p. 5

ΛCDM model[213] and, in general, several observations in "*these real voids are in gross conflict with the concordance* ΛCDM *model*."[214] However, according to the authors, these same data are compatible with a model that correlates the positive variation of the expansion velocity with the differential effects of the backreaction of the inhomogeneities of the Universe.[215]

In addition to taking seriously the backreaction, calculated by Buchert, by Li and Schwarz and by Hunt and Sarkar, Wiltshire also takes into account the effects of the differential running of time as measured by clocks in voids and walls. In 2008, Wiltshire published his own approximate results of the Buchert-Wiltshire model.[216] He showed that the deceleration of the expansion of the Universe is less in voids than in walls. The voids enhance the redshift of the light that passes through them, and since our galaxy is at the center of a huge void, we observe a nearby expansion deceleration that is less than the global-average. In general, the expansion deceleration in walls with time τ_w differs $5.5\,cm$ per s^2 ($5.5*10^{-10}\,m\,s^{-2}$) from the deceleration in voids with time τ_v. This seems little, but the accumulated effect through the entire history of the Universe, since the *Big Bang*, is large, that is, the lapse function is $1.42 < d\tau_v / d\tau_w < 1.46$ if we do not take into account the backreaction, and $d\tau_v / d\tau_w = 1.38^{+0.06}_{-0.03}$ if we do take it into account, which makes up for a difference of 38%.[217] This unequal deceleration in voids and walls implies that the Hubble parameter is not equal in different regions of the Universe.

I will first present the mathematical equation of the Hubble constant and its variation, considered by Wiltshire himself to be one of the most important equations of his model,[218] and then reproduce some of the estimates of its differential values.

MATHEMATICAL BOX 15. HUBBLE CONSTANTS IN WILTSHIRE'S MODEL

In mathematical box 14, we already came to know the subjacent, global-average Hubble constant:

$$H(t) = \frac{dt}{d\tau_w} H_w = \frac{dt}{d\tau_v} H_v \qquad (140)$$

where:

$$H_w \equiv \frac{1}{a_w}\frac{da_w}{dt} \text{ and (26 B) } H_v \equiv \frac{1}{a_v}\frac{da_v}{dt} \qquad (141)$$

Wiltshire defines the average and variation of the Hubble constant as a function of the

[213] Paul Hunt & Subir Sarkar, "Constraints on large scale inhomogeneities from WMAP-5 and SDSS: confrontation with recent observations", arXiv:0807.4508, Figure 5, p. 13
[214] Paul Hunt & Subir Sarkar, "Constraints on large scale inhomogeneities from WMAP-5 and SDSS: confrontation with recent observations", arXiv:0807.4508, p. 1
[215] Nan Li & Dominic Schwarz, "Scale dependence of cosmological backreaction", arXiv:astro-ph/0710.5073, p. 5
[216] David Wiltshire, "Cosmological equivalence principle and the weak field limit", in. *Physical Review D*, vol. 78 (2008) y "Exact Solution to the Averaging Problem in Cosmology", in: *Physical Review Letters*, vol.99 (2007)
[217] David Wiltshire, "Cosmological equivalence principle and the weak field limit", in. *Physical Review D*, vol. 78 (2008) y arXiv:0809.1183 (2008): 9
[218] Private communication of David Wiltshire to the author, April 1st, 2009

wall time τ_w and the global-average time t.[219] The first term of the right side of the equation represents the Hubble constant as a function of τ_w and the second term, the rate of change of the constant $(dt / d\tau_w)$.

$$H(\tau_w) = \frac{1}{a}\frac{da}{d\tau_w} = \frac{1}{\bar{a}}\frac{d\bar{a}}{d\tau_w} - \frac{d\tau_w}{dt}\frac{d\left(\frac{dt}{d\tau_w}\right)}{d\tau_w} \qquad (142)$$

Since, by definition:

$$\frac{1}{\bar{a}}\frac{d\bar{a}}{d\tau_w} = \frac{dt}{d\tau_w}\frac{1}{\bar{a}}\frac{d\bar{a}}{dt} = \frac{dt}{d\tau_w}\bar{H}(t) \qquad (143)$$

we obtain, combining (142)and (143):

$$H(\tau_w) = \frac{dt}{d\tau_w}\bar{H}(t) - \frac{d\left(\frac{dt}{d\tau_w}\right)}{dt} \text{ (note 220)} \qquad (144)$$

Equation (144) gives us the Hubble constant and its variation over time, as well as the rate of decrease of the lapse function, and in order to establish its value, one must make measurements with two clocks, that is the one with the global-average-time t and the other one, with the proper time of the observer in a galaxy cluster τ_w. The value of the Hubble constant is different, when measured with the same global-average clock in different regions, that is, in walls, in voids, or in the Universe at large; the estimates of its values in this case are $H_w(t) = 34.9$, $H_v(t) = 52.4$ y $H(t) = 48.2$, respectively. The present Hubble constant also varies, when measured in the same region, with different clocks, that is the wall-clock, that runs slower, and therefore yields a higher expansion velocity, or the global average clock, that runs faster, and therefore yields a lower velocity, resulting in $H_0(\tau_w) = 61.7\, kms^{-1}\, Mpc^{-1}$ and $H_0(t) = 48.2\, kms^{-1}\, Mpc^{-1}$, respectively, a difference of 28% (note 221).

The apparent acceleration of the expansion of the Universe.- Wiltshire also redefines the redshift, taking into account the difference between the observer's clock and time located in a wall, τ_w, and the global average time t (see next mathematical box).

---

[219] Wiltshire explained to me that he has omitted the suffix w in his article from equation (38) onwards and stressed the importance of equation (48): "*This equation relates the thing we interpret as the average Hubble parameter H to an underlying bare Hubble parameter $\bar{H}$. Both of these are "measurable." The point is that there is not only an average Hubble parameter, but a variance in the Hubble parameter, if referred to one set of clocks, such as ours. Equation (42) quantifies both the average and their variance.*"

[220] David Wiltshire, "Exact solution to the averaging problem in cosmology", arXiv:0709.0732 (2007): 2, equation 8 and "Cosmic clocks, cosmic variance and cosmic averages", in: *New Journal of Physics* (2007): 25, equation 42

[221] David Wiltshire, "Cosmological equivalence principle and the weak-field limit", in: *Physical Review D*, vol. 78 (2008): 12.

MATHEMATICAL BOX 16. THE REDSHIFT IN WILTSHIRE'S MODEL

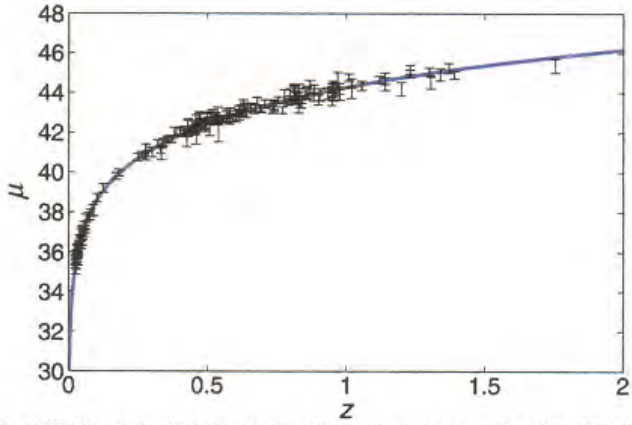
The redshift z determined by observers in dense regions is defined as a function of the redshift z determined by observers whose watch measures the global-average time:

$$1+z=\frac{a_0}{a}=\frac{\frac{dt}{d\tau_w}\bar{a}_0}{\frac{dt_0}{d\tau_{w_0}}\bar{a}}=\frac{dt/d\tau_w}{dt_0/d\tau_{w_0}}(1+z) \quad \text{and} \quad 1+z=\frac{dt_0/d\tau_{w_0}}{dt/d\tau_w}(1+z) \qquad (145)$$

The redshift of supernovae cannot easily be established, because of the dust in the host galaxy of supernovae and colour variations, but observers in dense regions observe larger redshifts of the same supernova then observers in voids or global-average ones.

In the Riess sample of 182 supernovae type 1a, the Wiltshire model is a *perfect fit* of the relation between distance and redshift, as can be appreciated in the following graph.[222] The same cannot be said for other supernovae samples. We will return to this point shortly.

Graph.- The relationship between distance and redshift in the Buchert-Wiltshire paradigm



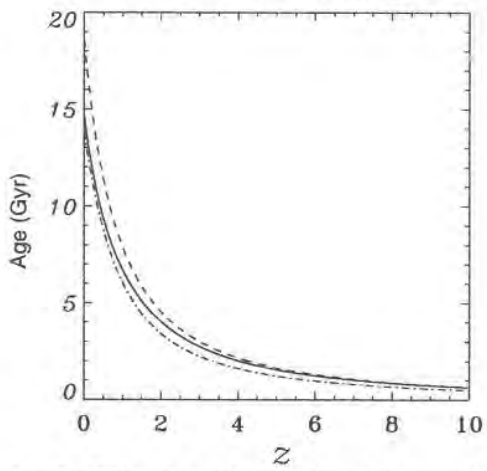
David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in: New Journal of Physics (2007): p. 37, figure 2

Explanation: Distance modulus, $\mu \equiv m - M = 5.\log_{10}(d_L) + 25$, versus redshift, z, with d_L in units Mpc. The theoretical curve for a FB model with $H_0 = 62.0\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\bar{\gamma}_0 = 1.38$, $f_{v0} = 0.759$ is compared to the 182 SneIa, excluding the 'Hubble bubble' points at $z \leqslant 0.023$ of the Riess *et al* Gold06 data set [51]. For these parameter values $\chi^2 = 163.2$, or 0.9 per degree of freedom.

The Buchert-Wiltshire model results in a new age of the universe (remember that $t_{age} = r/v = 1/H$), estimated to be a thousand million years older than is usually assumed in the standard ΛCDM model (see next graph).[223]

[222] David Wiltshire, *"Cosmic clocks, cosmic variance and cosmic averages"*, in *New Journal of Physics* (2007): 35
[223] *Ibidem*, p. 37

Graph.- The relationship between age of the Universe and redshift, according to ΛCDM and the new Buchert-Wiltshire paradigm



Source: David Wilshire, "Cosmic clocks, cosmic variance and cosmic averages", in: New Journal of Physics (2007): p. 37, figure 5

Explanation: The expansion age as seen by wall observers in galaxies for a FB model with $H_0 = 62.0\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\bar{\gamma}_0 = 1.38$, $f_{v0} = 0.759$ (solid line), is compared to the expansion age as seen by volume-average observers for the same parameters (dashed line). The expansion age for standard ΛCDM cosmology with the 'concordance' parameters $H_0 = 71.0\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_{M0} = 0.268$, $\Omega_{\Lambda 0} = 0.732$ (dot-dashed line) is also shown for comparison.

What is said until now prepares the way to resolve the question of the apparent acceleration of the expansion of the Universe, which is attributed by many cosmologists to the negative pressure of dark energy, and which Kolb, Matarrese and Riotto attribute to the sole influence of the backreaction. In Wiltshire's view, the backreaction by itself is too small to explain this apparent acceleration of the expansion velocity.[224]

¿Why do I speak of *apparent* acceleration? ¿Is the acceleration not real? Yes and no. The fundamental relativistic principle that guides us is the following one, formulated by Wiltshire: "*Systematically different results will be obtained when averages are referred to different clocks*."[225] If we observe a supernova, that moves away from us, and is located at the other end of the void that surrounds us, the velocity and redshift of its light, passing through this large void, and reaching an observer in a strong gravitational field, will be higher when measured by the observer's clock than by the clock in the void. The clock in the more dense region runs more slowly than the one in the void. For that reason, the deceleration measured with the wall clock will be different from the one measured with the clock in the void. When measured with the clock in the void, the supernova will appear to move away from us at a slower rate, and its light will appear to have a smaller redshift, then when measured with the wall clock, because of the differential clock rates. A terrestrial observer, measuring with his own, slower running wall clock the redshift of the supernova's light, might observe an acceleration, whereas an observer located in the void surrounding us, might measure, with his faster running clock, a deceleration "*it is quite possible to obtain regimes in which the wall observers measure apparent acceleration,*

[224] *Ibidem*, p. 22
[225] David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in: *New Journal of Physics* (2007): 27

$q < 0$, *even though void observers do not* $[q > 0]$."[226] Today, the parameter of the global deceleration has a value of $q = -0.0428$ (an apparent acceleration), while this same parameter, in the proper time of the global-average observer has a value of $\bar{q} = 0.0153$ (a deceleration).[227] This is why Einstein's theory is called general relativity: time and space are relative.

The history of the Universe has three periods: 1) a brief period that was dominated by energy; 2) a long period that is now reaching its end, dominated by matter; 3) and a third epoch that is now beginning, dominated by voids. Actually, right now we live in a transition period between the matter dominated epochs and the following one dominated by ever larger volumes of voids. It is precisely in such a transition period that we may observe an apparent acceleration:

"*Depending upon parameter values, it is possible for wall observers to register an apparent acceleration with the deceleration parameter (...) taking values of* $q < 0$. *Backreaction and the rate of decrease of* $\bar{\gamma}\left[= dt / d\tau_w\right]$ *are largest in an epoch during which the universe appears to undergo a void-dominance transition, or equivalently a transition in which spatial curvature* Ω_k *becomes significant. The reason for apparent acceleration at such an epoch (...): in the transition epoch the volume of the less rapidly decelerating regions increases dramatically, giving rise to apparent acceleration in the volume-average. We must be careful to note that these statements are true, when referred to one set of clocks, such as our own [with time* τ_w*]* (…). [*C*]*osmic 'acceleration' is an apparent effect, depending crucially on the position of the observer and local clocks. Both observers register a deceleration parameter close to zero, a general feature of a universe which undergoes a void-dominance transition. According to a wall observer in a galaxy, apparent acceleration begins at an epoch* $z = 0.909$ *for the present parameters, when the universe is 7.07 Gyr old, a little under half its current age. The void fraction at this epoch is* $f_v = 0.587$."[228]

The following mathematical box summarizes the numerical values of the most important parameters in the new, relativistic Buchert-Wiltshire model.

MATHEMATICAL BOX 17. THE VALUES OF SOME COSMOLOGICAL PARAMETERS IN WILTSHIRE'S RELATIVISTIC TIMESCAPE MODEL

Below, I reproduce estimates of the empirical values of different cosmological parameters in the relativistic Buchert-Wiltshire model.

The author (≡Wiltshire) has the following differential clock rates in *voids* and w*alls* (the *lapse* function):[229]

[226] David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", en *New Journal of Physics* (2007): 29
[227] David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", en *New Journal of Physics* (2007): 34
[228] David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", *New Journal of Physics* (2007): 30,35
[229] David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in: *New Journal of Physics* (2007): 21 (equation 28) and David Wiltshire, "Exact solution to the averaging problem in cosmology", arXiv:0709.0732 , (2007): 2 (equation 7)

$$\gamma_w = 1 + \frac{(1-h_r)}{h_r} f_v = \frac{dt}{d\tau_w} = 1.381^{+0.06}_{-0.05}, \text{ (note 230)} \quad (146)$$

Let us remember from box 15 that:

$$f_w(t) + f_v(t) = 1 \text{ (note 231)} \quad (147)$$

and the present day fraction of *voids* in the Universe is:(note 232)

$$f_v = 0.759^{+0.12}_{-0.09} \quad (148)$$

so that the fraction of walls is:

$$f_w \approx 0.241 \quad (149)$$

From (121), (146) and (148), we obtain:

$$h_r(t) = H_w / H_v = \bar{\gamma}_v / \bar{\gamma}_w = d\tau_w / d\tau_v = 0.666 \quad (150)$$

and from (121) and 146), we obtain:

$$\bar{\gamma}_v \cong \frac{dt}{d\tau_v} = 0.92 \quad (151)$$

Wiltshire also gives an estimate of the global-average Hubble constant, called the *bare* Hubble constant, since it has a bar on top of the H :[233]

$$\bar{H}(t) = 48.2^{+2.0}_{-2.4}\ km\, s^{-1}\ Mpc^{-1} \quad (152)$$

Let us remember that (119) and (122) give us:

$$\bar{H}(t) = \bar{\gamma}_w H_w = \bar{\gamma}_v H_v = f_w H_w + f_v H_v \quad (153)$$

so that we obtain, from (119), (122) and (152):

$$H_w(t) = \frac{1}{a_w}\frac{da_w}{dt} = 34.9 \quad (154)$$

$$\text{and} \quad H_v(t) = \frac{1}{a_v}\frac{da_v}{dt} = 52.4 \quad (155)$$

The author estimates the following value of the universal Hubble constant measured with the slower running clocks located in the walls, called the *dressed* Hubble constant. The fact that the wall clock runs more slowly, means that it registers the expansion velocity as higher than it would be if measured with the

[230] See David Wiltshire, "Cosmological equivalence principle and weak-field limit", in: *Physical Review D*, vol. 78 (2008): 9; Ben Leith, Cindy Ng & David Wiltshire, "Gravitational energy as dark energy: Concordance of cosmological tests", in: *Astrophysical Journal* vol. 672 (2007): 4

[231] Peter Smale & David Witshire, "Supernova tests of the timescape cosmology," en: arXiv:1009.5855v1, p.3

[232] David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in: *New Journal of Physics* (2007): 36-37 (figs. 4, 5) and Ben Leith, Cindy Ng & David Wiltshire, "Gravitational energy as dark energy: Concordance of cosmological tests", in: *Astrophysical Journal* vol. 672 (2007): 4

[233] David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in: *New Journal of Physics* (2007): 21; and in: "Cosmological equivalence principle and weak-field limit", in: *Physical review D*, vol. 78 (2008): 12 and Ben Leith, Cindy Ng & David 0, "Gravitational energy as dark energy: Concordance of cosmological tests", in: *Astrophysical Journal* vol. 672 (2007): 4

global-average clock, because it seems that the same distance is covered in less time.[234] Let us remember equation (144) of mathematical box 15:

$$H_0(\tau_w) = \frac{dt}{d\tau_w} H(t) - \frac{d\left(\frac{dt}{d\tau_w}\right)}{dt} = 61.7^{+1.2}_{-1.1}\ kms^{-1}\ Mpc^{-1} \qquad (156)$$

Let us also remember that Wiltshire proposes:[235]

$$Q = 6f_v(1-f_v)\frac{H^2(1-h_r)^2}{[h_r + f_v(1-h_r)]^2} = 6f_v(1-f_v)(H_v - H_w)^2\ (^{\text{nota 236}}) \qquad (157)$$

From (157), (148), (150) and (152), or from (157), (148), (154) and (155), we obtain:

$$Q \approx 336 \qquad (158)$$

Wiltshire proposes:[237]

$$\overline{\Omega}_Q = -\frac{Q}{6H^2} = -f_v(1-f_v)\left(\frac{(1-h_r)^2}{[h_r + f_v(1-h_r)]^2}\right) \qquad (159)$$

From (148), (150) and (159), we obtain:

$$\overline{\Omega}_Q \cong -0.0241 \qquad (160)$$

We can verify this result independently through another equation: [238]

$$Q = \frac{2\dot{f}_v^{\,2}}{3f_v(1-f_v)} \Rightarrow \qquad (161)$$

From (148), (158) and (161), we obtain:

$$\dot{f}_v^{\,2} = \frac{Q * 3f_v(1-f_v)}{2} = 92.1275 \qquad (162)$$

Since the author defines $\overline{\Omega}_Q$ also as:

---

[234] Ben Leith, Cindy Ng & David Wiltshire, "Gravitational Energy as Dark Energy: Concordance of Cosmological Tests", in: *The Astrophysical Journal*, vol. 672 (2008): Table I, p L94 and David Wiltshire, "Cosmological equivalence principle and weak-field limit", in: *Physical Review D*, vol. 78 (2008): 12

[235] David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in: *New Journal of Physics* (2007): 21 (equation 31)

[236] This is the same result as obtained by David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in: *New Journal of Physics* (2007): 21, equation (31)

[237] David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in: *New Journal of Physics* (2007): 9 (equation 10)

[238] David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in: *New Journal of Physics* (2007): 22 (equation 33)

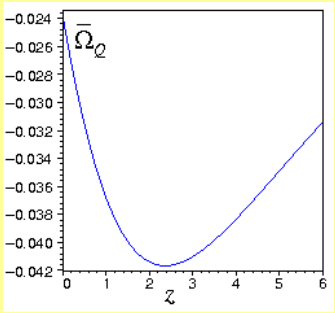
$$\overline{\Omega}_Q = \frac{-\dot{f}_v^{\,2}}{9 f_v (1-f_v) H^2} \text{ :(note 239)} \quad (163)$$

it follows, from (148), (152), (162) and (163), that:

$$\overline{\Omega}_Q \cong -0.0241 \quad (164)$$

The results of equation (164) and equation (160) are identical, Q.E.D. Wiltshire stresses the point that this is the present value of $\overline{\Omega}_Q$. The evolution of the value of $\overline{\Omega}_Q$ in time (represented by redshift z) can be appreciated in the following graph:[240]

Gráfica.- The value of $\overline{\Omega}_Q$ as a function of the redshift z



In a way analogous to the three values of the Hubble constant in the equation of the backreaction Q, there are three constants of curvature, one for the voids (k_v), one for the walls (k_w) and one global-average (k). Since we do not know the values of k_v and k_w, neither the values of a_v and a_w, we cannot resolve directly the following equations for Ω_k and $\langle R \rangle$:

$$\langle R \rangle_H = \frac{6k_v f_v}{a_v^{\,2}} + \frac{6k_w f_w}{a_w^{\,2}} \quad (165)$$

$$\Omega_k = -\frac{k_v f_v}{H^2 a_v^{\,2}} - \frac{k_w f_w}{H^2 a_w^{\,2}} = -\frac{\langle R \rangle}{6H_D^{\,2}} \quad (166)$$

But we can solve equation (113):

$$\Omega_M = \frac{8\pi G \langle \rho \rangle}{3H^2} \quad \Rightarrow \quad \Omega_M = 0.333765 \quad (167)$$

Since:[241]

---

[239] David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in: *New Journal of Physics* (2007): 25 (equation 45); and Peter Smale & David Witshire, "Supernova tests of the timescape cosmology," en: arXiv:1009.5855v1, p.4

[240] Courtesy of David Wiltshire in a private communication of May 5th, 2009

[241] David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in: *New Journal of Physics* (2007): 27 (equation 55) and David Wiltshire, "Exact solution to the averaging problem in cosmology", arXiv:0709.0732 , (2007): 2 (between equations 6 and 7)

$$\Omega_M = \bar{\gamma}_w^{\,3}\overline{\Omega}_M \qquad (168)$$

it follows, from (146) and (168), that the global-average matter density of the Universe has the following value:[242]

$$\overline{\Omega}_M = 0.125 \qquad (169)$$

Since it is true that: [243]

$$\overline{\Omega}_M + \overline{\Omega}_Q + \overline{\Omega}_k = \frac{8\pi G\langle\rho\rangle}{3H^2} + \frac{-Q}{6H^2} + \frac{-\langle R\rangle}{6H^2} = 1 \Rightarrow \overline{\Omega}_k = 1 - \overline{\Omega}_M - \overline{\Omega}_Q \qquad (170)$$

so it follows from (164), (169) and (170) that:[244]

$$\overline{\Omega}_k = 1 - 0.125 - (-0.024) \cong 0.899 \qquad (171)$$

The non-zero value of $\overline{\Omega}_k$ implies that the global-average Universe is open ($k < 1 \Rightarrow \Omega_k > 1 \;\Rightarrow \overline{\Omega}_M + \overline{\Omega}_Q = 1 - \overline{\Omega}_k < 1$). The author also obtains the *negative* value of the effective deceleration parameter, as measured with wall clocks (the *dressed* parameter), revealing an *apparent* acceleration of the Universe:[245]

$$q(\tau_w) = \frac{-(1-f_v)(8f_v^{\,3} + 39f_v^{\,2} - 12f_v - 8)}{(4 + f_v + 4f_v^{\,2})^2} \qquad (172)$$

the empirical value of which, can be obtained from (148) and (172):

$$q(\tau_w) = -0.0428^{+0.012}_{-0.0002} \qquad (173)$$

However, the value of the deceleration parameter, as measured by global-average clocks is *positive*. That means that, *in fact, the expansion of the Universe is decelerating*:[246]

$$q(t) = \frac{1}{2}\Omega_M + 2\Omega_Q = \frac{1}{2}\Omega_M - 2\frac{f_v(1-f_v)(1-h_r)^2}{[h_r + f_v(1-h_r]^2} \qquad (174)$$

And through (148), (150), (169) and (174), we obtain:

$$\bar{q}(t) = 0.015348 \qquad (175)$$

And from equation (174) we can also conclude, that there were two moments in the history of the Universe when the deceleration parameter $\bar{q}(t) = \frac{1}{2}\overline{\Omega}_M$, that is to say, when $\overline{\Omega}_Q = 0$, in the early history of the Universe, when $h_r \to 1$, and at the

---

[242] David Wiltshire, "Exact solution to the averaging problem in cosmology", arXiv:0709.0732 , (2007): 42 and Ben Leith, Cindy Ng & David Wiltshire, "Gravitational energy as dark energy: Concordance of cosmological tests", in: *Astrophysical Journal* vol. 672 (2007): 4

[243] David Wiltshire, *"Cosmic clocks, cosmic variance and cosmic averages*", in: *New Journal of Physics* (2007): 9 (equation 10)

[244] Value confirmed by David Wiltshire in a private communication of May 5th, 2009

[245] David Wiltshire, "Exact solution to the averaging problem in cosmology", arXiv:0709.0732 (2007): 4 (equation 26)

[246] David Wiltshire, *"Cosmic clocks, cosmic variance and cosmic averages*", in *New Journal of Physics* (2007): 28, equation (61)

final moments of the history of the Universe, when the entire Universe is an immense void $f_v \to 1$.[247]

In February 2009, Kwan, Francis and Lewis compared the capacity of both the standard ΛCDM model of a flat universe and the relativistic time-scape model of an open universe (according to Wiltshire and Leich), to be compatible with assumptions about the matter density parameter of the universe Ω_M and to explain observations of redshift of supernovae SNe 1a. They concluded, on the basis of Bayesian probability, that the ΛCDM model is a better fit for the Union and Constitution datasets than the TS (=timescape) model (initially known as the FB (=fractal bubble) model), as one can appreciate in the following graph.

Graph.- Comparison of the estimates of Ω_M in the flat ΛCDM model and the relativistic time-scape model [248]

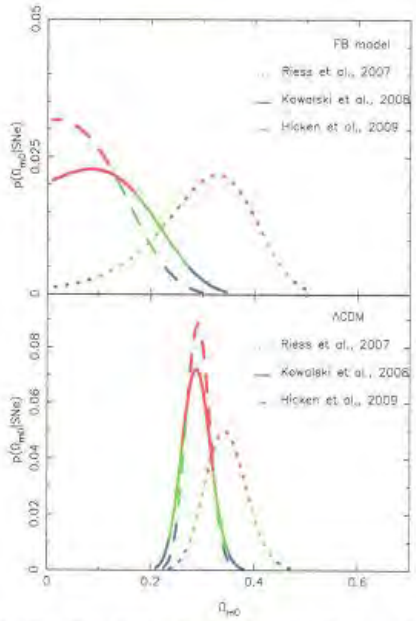


Figure 1. *Top*: Marginalised posterior distributions for the FB model using the Union compilation (Kowalski et al. 2008) in solid lines, the Gold SNe Ia sample from Riess et al. (2007) in dotted lines, and the Constitution set (Hicken et al. 2009) in dashed lines. The colours correspond to the confidence limits, red is 1σ, green is 2σ and blue is 3σ. *Bottom*: As above except using ΛCDM.

Smale and Wiltshire answered to these criticisms in September 2010. They showed, among other things, that

1) Kwan and his team fail to exclude data below the scale of statistical homogeneity, where things are really very inhomogeneous, using even at this scale the assumption of the flat global-average parameter values. Adjusting for this mistake, the Ω_{m0} values, however,

[247] David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in: *New Journal of Physics* (2007): 28
[248] Juliana Kwan, Matthew Francis & Geraint Lewis, "Fractal Bubble Cosmology: A concordant cosmological model?", arXiv:0902.4249 (2009): 2

suffer only a slight correction, not enough to explain why the TS model seems to be such a bad fit for the Union and Constitution samples.

2) Once the suitable cuts are made, one must also take into account the different data reduction methods used in different datasets. If the SALT/SALT II fitter is used, as is the case in four supernova samples (the Union sample, the Constitution sample, the Salt2 sample, and the Union 2 sample), then the Bayesian evidence favors the flat ΛCDM model over the TS model. But if the MLCS2k2 fitter is used, as is the case in four other samples (the Riess07 sample, the MCLS17 sample, the MLCS31 sample and the SDSS-II sample), then "*Bayesian evidence favours the TS model over the spatially flat* ΛCDM *model*."[249] This means that the "*primary question is the method of data reduction.*"[250]
3) Even in the case that the Bayesian evidence favours one model over the other, the difference is so slight that it is inconclusive according to Roberto Trotta's version of Jeffreys' scale of Bayesian probability.[251]

The following table gives some of the evidence produced by Smale and Wiltshire

Sample (dataset)	Data reduction method	Sample size N	χ^2	Ω_{m0} expectation value	f_{v0}	χ^2 corrected	Ω_{m0} corrected
Union TS	**S**	**307**	**320**	**0.12**	**0.91**	**351**	**0.13**
Union ΛCDM	**S**	**307**	**311**	**0.29**	**/**	**344**	**0.28**
Const. TS	**S**	**397**	**471**	**0.10**	**0.93**	**320**	**0.13**
Const. ΛCDM	**S**	**397**	**/**	**0.28**	**/**	**313**	**0.29**
SALT2 TS	S2	352	347	0.11	0.92		
Union2 TS	**S2**	**557**	**551**	**0.08**	**0.95**		
Union2 ΛCDM	**S2**	**557**	**/**	**0.274**	**/**		
Riess07 TS	**M**	**182**	**163**	**0.29**	**0.79**		
MLCS17 TS	M	372	403	0.18	0.87		
MLCS31 TS	M	366	433	0.07	0.95		
SDSS-II TS	**M**	**288**	**241**	**0.38**	**0.72**		
SDSS-II ΛCDM	**M**	**288**	**238**	**0.31**	**/**		

[249] Peter Smale & David Witshire, "Supernova tests of the timescape cosmology," en: arXiv:1009.5855v1, p. 18
[250] Peter Smale & David Witshire, "Supernova tests of the timescape cosmology," en: arXiv:1009.5855v1, p. 10
[251] See Mathematical BOX 8, above.

Let us now compare the Bayesian evidence fore the ΛCDM and *TS* models in the case of the Riess07, Union and Constitution samples, but taking into account only the 140 supernovae that these samples have in common. In all four cases, according to Smale and Wiltshire, "*we replaced the spatially flat ΛCDM luminosity distance by one that computes the TS luminosity distance, leaving the rest of the simple_cosfitter code un changed.*"[252] The authors say that "*the Bayes factors [are] representing the integrated likelihood of the TS model over the spatially flat ΛCDM model.*" (note 253) In these three cases the natural logarithm is $\ln B = 0.14$, $\ln B = 0.14$ and $\ln B = 0.17$, respectively, representing a 15% to 19% difference in the likelihood of not being false between the ΛCDM and *TS* models. The authors conclude that "*these results indicate that the models are statistically indistinguisable for the 140 SNe 1a regardless of the fitter used*" and "*both models are a very good fit.*"[254]

In 2007, Mustapha Ishak and Roberto Sussman demonstrated exactly the same thing for another relativistic model. The observational data of the supernovae 1a are concordant with both the flat ΛCDM model which is Newtonian and homogeneous, and has a component of dark energy, and the Szekeres model, which is relativistic and inhomogeneous, and has no dark energy component.[255]

Graph.- The supernovae type 1a data interpreted in the ΛCDM and relativistic models

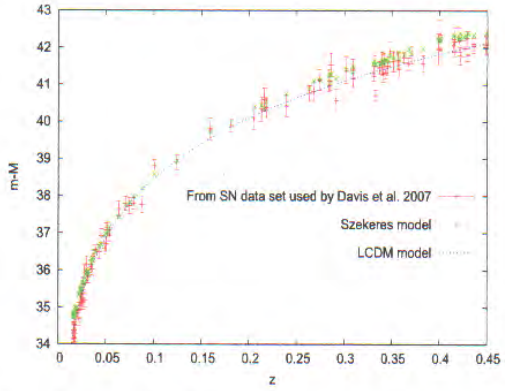


FIG. 1: Supernova fits for the Szekeres model (green crosses) and LCDM (blue curve) models. The data is 94 Supernova (up to $1 + z = 1.449$) from Davis et al 2007, Wood-Vasey et al 2007, and Riess et al 2007 [57, 58, 59]. The Szekeres model fits the data with a $\chi^2 = 112$. This is close to the $\chi^2 = 105$ of the LCDM concordance model. Because of the possible systematic uncertainties in the supernova data, it is not clear that the difference between the two χ^2 is significant. Furthermore, other Szekeres models remain to be explored. The Szekeres model used is also consistent with the requirement of spatial flatness at CMB scales.

[252] Peter Smale & David Witshire, "Supernova tests of the timescape cosmology," en: arXiv:1009.5855v1, p. 11
[253] Peter Smale & David Witshire, "Supernova tests of the timescape cosmology," en: arXiv:1009.5855v1, p. 12
[254] Peter Smale & David Witshire, "Supernova tests of the timescape cosmology," en: arXiv:1009.5855v1, p. 12, 18, my underlining
[255] Mustapha Ishak, Roberto Sussman *et al.*, "Dark Energy or Apparent Acceleration Due to a Relativistic Cosmological Model More Complex than FLRW?". arXiv:0708.2943 (2008)

Two models explain the same observational data. Both explain the apparent acceleration of the expansion of the Universe. From the point of view of Bayesian probability calculus the same data corroborate both models,[256] the ΛCDM model with $\Omega_\Lambda = 0.73$ ($\chi^2 = 105$) having a slightly higher probability than the relativistic Szekeres model with $\Omega_\Lambda = 0$ ($\chi^2 = 112$). However, "*in view of the possible uncertainties involved in the supernova data, it is not clear that the difference between the two* χ^2 *is significant*."[257]

If two different theoretical models explain the same phenomena, we have to apply Ockham's raisor and give preference to the more simple, and physically orthodox model, which is the relativistic one.

The evolution of the large structures in the Universe.- We just saw that there exist considerable variations of the value of the Hubble constant, depending on the clock that is used for the measurement. If we do not take into account these variations, the models used to simulate the evolution of large structures in the Universe are mistaken and, as a consequence, "*all steps in measuring masses of galaxy clusters would need to be carefully reconsidered*."[258] In the beginning, when the Universe was almost perfectly homogeneous, and the perturbations were minimal, a critical density of uniform matter existed that determined whether the Universe would eventually collapse or not., and which regions would collapse gravitationally to form large structures, and which would not, eventually creating the large. To investigate the evolution of large structures, it is necessary to have an estimate of this background critical density. Wiltshire comments that in the standard model, a fundamental mistake may be made in the way this estimate is obtained:

"*I will now make the following crucial physical observation. By the evidence of the CMBR, the universe at last scattering was very close to being truly homogeneous and isotropic. Therefore, and operational definition of critical density does exist, provided we assume the Copernican Principle and accept that the universe was globally smooth at that epoch and not just in our present past horizon volume. (..)*

At the epoch of last scattering, t_i*, the Hubble expansion was uniform, as the local velocity perturbations were tiny. Given a uniform initial expansion rate there must have existed a uniform critical density of matter required for gravity to be able to eventually bring that expansion to zero. This critical density,* $\rho_{cr}(t_i)$*, therefore sets a universal scale which delineates the boundary between density perturbations which will become bound, as opposed to density perturbations which are unbound.*

This may seem a trivial point. However, when one considers an inhomogeneous evolution, it is clear that the average Hubble parameter on a given domain does not correspond to the time evolved critical density, whether in a dressed or undressed form. The naive use of the Friedmann equation in cosmology to date means that we could well be making a gross error in the choice of background in structure formation studies. In particular, we estimate the critical density by

---

[256] See mathematical BOX 8 of Section 15.2 of John Auping, *El Origen y la Evolución del Universo* (2009)

[257] Mustapha Ishak, Roberto Sussman *et al*., "Dark Energy or Apparent Acceleration Due to a Relativistic Cosmological Model More Complex than FLRW?". arXiv:0708.2943 (2008): 5

[258] David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in *New Journal of Physics* (2007): 31

extrapolating back in time using our present Hubble parameter, H_0, assuming the evolution of the universe is smooth and that $3H_0^2/(8\pi G)$ is the critical density at the present epoch, when it is not. If the total density of the Universe is very close to one, this has the effect that we can miss-estimate the background density of the universe at early epochs where structure formation boundary conditions are set. We are in effect perturbing about the wrong background by implicitly assuming that the average density at the present epoch, as determined by the recent past in our past light cone, is identical to the universe as a whole. By cosmic variance there is no reason to expect this to be the case."[259]

As a consequence, "*neither the volume-average density parameter, $\overline{\Omega}_M$, nor the 'true' density parameter [in walls], Ω_{MW}, will take a numerical value close to those of the 'concordance' dark-energy cosmology.*"[260] The fact that in the relativistic Buchert-Wiltshire model, the Universe has an age that is a thousand million years larger than the 13.7 thousand million years normally estimated in the ΛCDM model, also gives more time to the formation of large structures.

Let us see this point with some detail. To analyze the distribution of the baryonic mass density of galaxy clusters, cosmologists generally use the Navarro-Frenk-White (= *NFW*) model.[261] The density profile in this model contains, among other factors, the critical density ρ_{cri}, which is defined as a function of the global-average Hubble parameter value at late epochs of the evolution of the Universe, and this value is very model-dependent. In the Buchert-Wiltshire model, its value is 40 to 80% of the value obtained in the *NFW* theorem.[262] This error margin is not trivial and is a consequence of the fact that the *NFW* model generates computer simulations in the context of Newtonian gravitational dynamics of N bodies, which yield results that are different from those obtained in a relativistic inhomogeneous model, as Wiltshire points out:

"*Physically, as long as ρ_{cri} represents a closure density, then it must correspond to the true critical density, $3\overline{H}^2/(8\pi G)$, and as indicated earlier this may be typically 40-80% of the value of the critical density estimated from the global average Hubble constant at late epochs. Unfortunately, the NFW model is essentially an empirical fit to the results of N-body CDM simulations in Newtonian gravity. Thus one cannot make any simple qualitative statement about how its use might change if the closure density is to be recalibrated. If similar empirical fits apply, then the only obvious deduction we can make is that if [the critical density] ρ_{cri} is effectively overestimated at late epochs, then the density contrast δ_C is effectively underestimated. Consequently, (...) we would expect the density contrast in galaxy clusters to be*

---

[259] David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in *New Journal of Physics* (2007): 10 (my underlining)
[260] David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in *New Journal of Physics* (2007): 27
[261] See Julio Navarro, Carlos Frenk & Simon White, "The Structure of Cold Dark Matter Halos", in: *The Astrophysical Journal*, vol. 462 (1996): 563-575 y "A Universal Density Profile from Hierarchical Clustering", in: *The Astrophysical Journal*, vol. 490 (1997): 493-508
[262] David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in *New Journal of Physics* (2007): 52

higher than is usually assumed."[263] The following mathematical box explains some aspects of the *NFW* theorem.

MATHEMATICAL BOX 18. THE MASS DENSITY OF GALAXY CLUSTERS

The total galaxy cluster mass is estimated by means of a Newtonian galaxy cluster mass function (see box 2). The mass density profile in the Navarro-Frenk-White theorem has its origin in the analysis of globular clusters and was later extended to galaxy clusters and shows us how baryonic mass is distributed in a certain spherical volume. The galaxy cluster density profile of Navarro-Frenk-White gives us the mass density in a spherical volume as a function of the critical density ρ_{cri}, the density contrast δ_C , and the radius r of the sphere:

$$\rho(r) = \rho_{cri} \frac{\delta_C}{(r/r_S)(1+r/r_S)^2} \quad (176)$$

The density contrast δ_C tells us how much more dense is the mass in the cluster center than in its outer regions:

$$\delta_C = \frac{200}{3} \frac{C^3}{[\ln(1+C) - C/(1+C)]} (^{\text{nota } 264}) \quad (177)$$

There are two empirical parameters, that is to say, the scale radius r_S, that is defined as $(r/r_S) = 1$ and the concentration parameter C, that is defined as the rate of radius r_{200} and the scale radius r_S :

$$C = r_{200}/r_S (^{\text{nota } 265}) \quad (178)$$

where the radius r_{200} is the radius at which the mass density is 200 times the average background radius, which is fixed according to the model that lies at the basis of the simulations. The following graph visualizes the various factors that intervene in the mass density profile.

The problem is that in this Newtonian model the critical density ρ_{cri} is defined as a function of the global-average Hubble parameter H and its corresponding redshift z :

$$\rho_{cri} = 3H^2(z)/(8\pi G) \quad (179)$$

whereas the 'true' critical density in the relativistic Buchert-Wiltshire model is:

---

[263] David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in *New Journal of Physics* (2007): 52

[264] Wiltshire transcribes the *NFW* equation with a mistake, that is: $\delta_C = 200C^3/[3\ln(1+C) - C/(1+C)]$., as can be seen from comparing David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in *New Journal of Physics* (2007): 52 with Julio Navarro, Carlos Frenk & Simon White, "The Structure of Cold Dark Matter Halos", in: *The Astrophysical Journal*, vol. 462 (1996): equation (4), p. 566

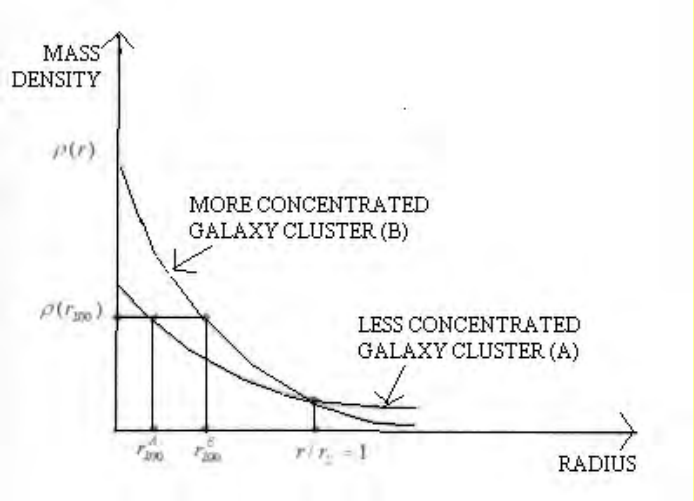
[265] Wiltshire transcribes the *NFW* equation with an annotation error, that is $C = r_{200}/r$. Compare David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in *New Journal of Physics* (2007): 52 with Julio Navarro, Carlos Frenk & Simon White, "The Structure of Cold Dark Matter Halos", in: *The Astrophysical Journal*, vol. 462 (1996), p. 566

$$\rho_{cri} = 3H^2(z)/(8\pi G) \qquad (180)$$

and the 'true' redshift, according to equation (145) in box 16 is:

$$1+z = \frac{dt_0/d\tau_{w_0}}{dt/d\tau_w}(1+z) \qquad (181)$$

Graph.- The galaxy cluster density profile of Navarro-Frenk-White



This is not a trivial error, because ρ_{cri} should correspond to the true critical density, $3H^2/(8\pi G)$ (equation 180 of box 18), which might be 40-80% of the value of the estimate derived from the Newtonian model density, $3H^2(z)/(8\pi G)$. This big error margin in the critical density profile of galaxy clusters in the models that suppose Newtonian dynamics explains several phenomena that would require the dark energy hypothesis in Newtonian models, for example:

1) From the supposed missing mass, Vikhlinin and his team, conclude that the dark energy hypothesis of thee ΛCDM model is necessary.[266] They point out that if the $\Omega_M = 0.25$, $\Omega_\Lambda = 0$, $h = 0.72$, would be valid, there appears to be a deficit of about 50% of massive galaxies, in the nearby range of $0.55 < z < 0.90$. If we abandon, however, the Newtonian assumptions of this computer model, the mass density in the sub sample of 11 nearby massive galaxy clusters might actually be sufficient.
2) We analyzed above the bullet cluster, and its supposed missing mass, that seems to validate the hypothesis about cold dark matter. Wiltshire comments this case: "*If there are any testable consequences that follow from such considerations [on the background mass density and the Hubble parameter], then they may possibly apply to highly dynamical non-equilibrium circumstances, such as that of the collision of the galaxy clusters observed in the bullet cluster 1E0657-558. The high velocity of the gas shock front trailing the smaller sub-cluster 1E0657-56 appears anomalously high as compared to expectations from the masses of sub-clusters inferred by using weak gravitational*

[266] See graph in Part 6.3

> *lensing and with the NFW or King profiles. Others*[267] *have argued that such high velocities may not be all that rare statistically speaking. Since it is quite possible that the dynamics of such systems may change in the present [relativistic] model, without invoking new forces of nature, this issue deserves further investigation.*"[268]

The anisotropies of the Cosmic Microwave Background Radiation *CMBR*.- We may speculate that there is a direct relationship between the evolution of large structures in the Universe and the conduct of the Cosmic Microwave Background Radiation. The graph of the *CMBR* anisotropies that I reproduced above[269] has five peaks. Wiltshire formulates a hypothesis about the various structures of the Universe corresponding to these peaks.[270] The enormous $100h^{-1}\,Mpc$ scale of statistical homogeneity would correspond to the first Doppler peak, which implies that there are no structures in the Universe in excess of this scale. The scale of the $30h^{-1}\,Mpc$ dominant voids would correspond to the second Doppler peak, namely, the first rarefaction peak, which is well within the nonlinear regime of structure formation. The third Doppler peak, which is the first compression peak within the nonlinear regime, would give the scale of the largest bound structures that have broken from the Hubble flow, namely galaxy clusters, and the fourth Doppler peak, which represents the second rarefaction peak, may possibly give an independent scale corresponding to mini-voids. Wiltshire concludes that "*these qualitative speculations about the correspondence of the Doppler peaks to the observed scales of present epoch structures should be verified from a numerical model of structure formation*."[271]

The most important point to be made here is not, however, the correspondence of various structures at different scales to the different Doppler peaks of the *CMBR*, but the fact that the wavelength and temperature of this radiation obtain different values when measured in different regions of the Universe. This is not a matter of putting in doubt the cosmographic observations of the *CMBR*, its temperature and its anisotropies, but rather a matter of interpreting these data from the point of view of different cosmological theories and models. We already encountered the recalibration of the redshift of electromagnetic waves,[272] among them the *CMBR*. There are important implications of recalibrating the parameters of the theoretical models used to interpret the data. Wiltshire explains that modern cosmology is in a habit of using the *CMBR* of gravitationally bound systems to obtain estimates of this radiation in early epochs of the Universe, and that this procedure might be wrong:

"*[T]he redshift of the CMBR as measured at a co-moving volume-average position is greater then in [gravitationally] bound systems. Since the volume-average CMBR temperature is used to calibrate several parameters associated with the primordial plasma, we need to recalibrate all quantities associated with the early universe. (...) An important consequence of the variation in*

---

[267] Among them Simon White of the *NFW* theorem

[268] David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in: *New Journal of Physics* (2007): 52 (my underlining)

[269] See graph Part 6.4

[270] David Wiltshire, "Cosmological equivalence principle and the weak field limit," in: *Physical Review D,* vol. 78 (2008): 11

[271] David Wiltshire, "Cosmological equivalence principle and the weak field limit," en: *Physical Review D,* vol. 78 (2008): 11

[272] See equation (50) of mathematical BOX 16

clock rates between ideal co-moving observers in galaxies and voids, is that the temperature of the CMBR will be lower when measured at the volume-average in voids, taking a value $T = \bar{\gamma}^{-1}T$ *, at any epoch, where* T *is the [apparently higher] temperature of the CMBR as seen by wall observers [like ourselves].*"[273]

We can make an estimate of this difference in temperature of the *CMBR* as measured from different locations and with different clocks. The temperature that we measure from our terrestrial position is $T_0 = 2.725\,K$, but, when measured from a global-average position, it is $T_0 = \frac{d\tau_w}{dt}T_0 = \frac{1}{1.38}(2.725)\,K = 1.975\,K$. All this has also consequences for the measurement of the anisotropy scale of the *CMBR:*

"[T]he fact that we observe an almost isotropic CMBR means that other observers should also measure an almost isotropic CMBR. However, it does not demand that such observers measure the same mean CMBR temperature as we, nor the same angular scale for the Doppler peaks in the anisotropy spectrum. Significant differences can arise due to gradients in the spatial curvature and associated gravitational energy. (...) [T]he locally defined or bare Hubble parameter H*, can be uniform even though voids appear to expand faster than the bubble walls which surround them, since cosmic clocks within voids tick faster on account of gravitational energy differences. Since our cosmological observations involve photons exchanged with objects in bound systems, we do not observe clocks in freely expanding space directly. Nevertheless, an ideal observer within a void would measure a somewhat older age of the Universe, and an isotropic CMBR with a lower mean temperature and an angular anisotropy scale shifted to smaller angles.*"[274]

Especially, we cannot assume a value of z_{dec} presumed by the *Wilkinson Microwave Anisotropy Probe,* because these values imply uncertainties and calibrations dependent on the standard ΛCDM model. To get from the calibrations of the algorithms that explain the Doppler peaks of the *CMBR* anisotropy specter, to the new algorithms implied by the relativistic Buchert-Wiltshire model, will require an enormous effort:

"*Whereas the underlying physics at the epochs of primordial nucleo-synthesis and recombination, when* $\bar{\gamma} \cong 1$*, is no different than usual, the fat that volume-average observers should measure a mean CMBR temperature of* $2.725\,\bar{\gamma}_0^{-1}$ *at the present epoch will affect all the usual calibrations of radiation-dominated era parameters inferred relative to present epoch co-moving observers. We are faced with the task of systematically re-deriving all the standard*

---

[273] David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in: *New Journal of Physics* (2007): 32, 40.

[274] David Wiltshire, "Exact solution to the averaging problem in cosmology", arXiv:0709.0732, ps. 1-2 (my underlining).

textbook calculations[275]*associated with the hot big bang, and making recalibrations where necessary.*"[276]

Specifically, some anomalies in the measurement of the large angle *CMBR* multipoles, might be explained by Wiltshire's new model, and the recalibration of the standard parameters will require an enormous computational effort:

""*[A] staitstical study of several possible systematic errors in the WMAP data by Freeman* et al *indicates that, of the several effects they studied, a 1-2% increase in the magnitude of the peculiar velocity attributed to the CMBR dipole was the only one which may potentially resolve anomalies associated with large angle multipoles. This is precisely the order of magnitude of effect we would expect from a* $|\Delta T|/T \approx 10^{-5}$ *contribution from a Rees-Sciama dipole. Effectively, our estimate of the magnitude of the peculiar velocity would imply a 1-2% systematic error [in the standard model] due to a small anomalous boost. Disentangling the small Rees-Sciama dipole from the dominant contribution of our own peculiar velocity with respect to the cosmic rest frame would require an enormous computational effort, and the sky maps would have to be redrawn. However, in the interest of our fundamental understanding of the universe, these steps should be taken.*"[277]

This recalibration of parameters of the cosmic rest frame would probably make the dark energy hypothesis superfluous, as Wiltshire points out: "*It is these recalibrations, which account for quasi local energy variations, which will allow us to obtain a viable model of the universe without dark energy.*"[278] The same point is made by Hunt and Sarkar: "*With the smaller global Hubble parameter, the* $WMAP-5$ *data on cosmic microwave background anisotropies can be fitted without requiring dark energy.*"[279]

7.4.- Conclusion

The ***conclusion*** of all this is that it may be possible to have a new, relativistic, inhomogeneous cosmological model, that can explain, among other things, the apparent acceleration of the expansion velocity of the Universe; the evolution of galaxy clusters; and the anisotropy scale of the *CMBR*, without any dark energy:

"*As theoretical physicists, we are altogether too much inclined to add all sorts of terms to the gravitational action (...), rather than thinking deeply about the basic operational issues of our subject. I believe we should guard the principles that have worked until they can be proven to fail. It is my own view that Einstein was correct about general relativity, and what I have presented here follows logically from his theory when combined with initial conditions.*"[280]

---

[275] Wiltshire mentions specifically Jim Peebles, *Principles of Physical Cosmology* (1993) and Edward Kolb & Michael Turner, *The Early Universe* (1990) and we could add Malcolm Longair, *The Cosmic Century* (2006) and Steven Weinberg, *Cosmology* (2008), among other works

[276] David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", en *New Journal of Physics* (2007): 40

[277] David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", en *New Journal of Physics* (2007): 47

[278] David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in: *New Journal of Physics* (2007): 32

[279] Paul Hunt & Subir Sarkar, "Constraints on large scale inhomogeneities from WMAP-5 and SDSS: confrontation with recent observations", arXiv:0807.4508, pág.1

[280] David Wiltshire, "Cosmic clocks, cosmic variance and cosmic averages", in *New Journal of Physics* (2007): 62

Not all cosmologists, at present, agree. For example, in his new book on cosmology, Steven Weinberg sustains the standard ΛCDM model and does not enter into discussion with the new relativistic model.[281] We do not find one single reference in his book to the work of Cooperstock and Tieu, Brownstein and Moffat, or Buchert and Wiltshire. In his book, *The Structure of Scientific Revolutions*, Kuhn explained that scientific revolutions take time in order to be accepted in the academic community.[282]

[281] Steven Weinberg, *Cosmology* (2009).
[282] Thomas Kuhn, *The Structure of Scientific Revolutions* (1996)

Relativistic gravitational dynamics and the rotation curves of galaxies

F. I. Cooperstock[1]* and S. Tieu[2]

*Department of Physics and Astronomy, University of Victoria,
P.O. Box 3055, Victoria, B.C. V8W 3P6 (Canada)

Abstract. We demonstrate that general relativity, universally accepted as the best theory of gravity, has important consequences for systems hitherto believed to be adequately treated with Newtonian theory. As an important example, we apply general relativity to the description of the dynamics of a galaxy. We show that the presently available data of the observed flattened galactic rotation curves can be accounted for without the currently conjectured vast stores of dark matter in galactic halos. The galaxy is modeled as a stationary axially symmetric pressure-free fluid. In spite of the weak gravitational fields and the non-relativistic source velocities, the mathematical system is still seen to be nonlinear. We determine that the mass density for the luminous threshold as tracked in the radial direction is approximately $10^{-21.75}$ kg·m^{-3} already for six galaxies studied thus-far and conjecture that this concordance will persist for galaxies yet to be analyzed. This threshold density has potential value for the understanding of galactic evolution. We present a velocity dispersion test to determine the extent, if of any significance, of matter that may lie beyond the visible/HI region. This is determined by examining the rotation curves at different galactic latitudes, bringing into consideration the global dynamical structure of the galaxy. The demand for global consistency applies not only to our own but also to all proposed models and theories.

INTRODUCTION

It is the generally held belief that Einstein's theory of gravity is primarily a theory for very strong gravitational fields and that it invariably provides only very small corrections to Newtonian predictions for weak fields. We have always resisted falling into this mindset mold. After all, it is well-understood that even very non-relativistic sources such as ordinary spinning rods or linear oscillators with their accompanying very weak gravitational fields, will emit gravitational waves yet there are no such waves at all within the context of Newtonian gravity. General relativity is much richer and much more complex than Newtonian gravity and it behooves us to avoid hasty judgments.

A prime example of the richness of general relativity concerns the dynamics of galaxies, vast expanses of billions of stars, for the most part moving in circles about a central axis of rotation in their generally weak gravitational fields. The motivation for the work which we will describe stems from the need to account for the observed essentially flat velocity rotation curves for galaxies. By "rotation curves" we mean the plots of the observed stellar velocities as a function of distance from the rotation axis. That these

[1] cooperst@uvic.ca
[2] stieu@uvic.ca

curves do not fall off with distance as would be expected on the basis of Newtonian gravity has been a central issue in astrophysics. There has been much speculation over the question of the nature of the "dark matter" that is generally believed to be required for the consistency of the observations of higher-than-expected stellar velocities within the context of Newtonian gravitational theory. While various researchers are now turning to gravitational lensing in the search for evidence for dark matter, probably the majority of researchers regard the flat galactic rotation curves as the key indicators. Clearly the issue is of paramount importance given that the dark matter is said to comprise the dominant constituent of an extended galactic mass [6]. The dark matter enigma has influenced particle theorists to devise acceptable candidates for the constitution of dark matter. At the present time, various researchers are hoping to discover dark matter particles in the LHC (Large Hadron Collider) experiments at CERN in Geneva. While physicists and astrophysicists have pondered the dark matter issue, other researchers have devised new theories of gravity to account for the observations (see for example [7, 8, 9, 10]). However the latter approaches, however imaginative, have met with understandable skepticism, as they have been devised solely for the purpose of the task at hand. To our knowledge, general relativity has never previously been proposed as a possible means of accounting for the flat rotation curves without invoking large stores of dark matter. However, general relativity remains the preferred theory of gravity with Newtonian theory as its limit where appropriate. General relativity has been successful in every test that it has encountered, going beyond Newtonian theory where required. Therefore, should it actually transcend Newtonian gravity in resolving the problem at hand, it would greatly alter the understanding of some basic aspects in physics. In what follows, we will set out to show that this is the case. We now provide an overview of the issues.

It is understandable that the conventional gravity approach has focussed upon Newtonian theory in the study of galactic dynamics since the galactic field is weak (apart from the deep core regions where black holes are said to be harboured, at least in some galaxies) and the stellar motions are non-relativistic ($v \ll c$). It was this approach that led to the inconsistency between the theoretical Newtonian-based predictions and the observations of the visible sources alone. To reconcile the theory with the observations, researchers subsequently concluded that to realize the observed motions, much more matter than that which was observed must be present to drive the high-velocity motions. Thus came the notion that a kind of dark matter must be present around galaxies in various cases massive halos that constitute the bulk of the extended galactic masses. While this might at first sight appear to be a relatively simple cure to the problem of motion, these massive halos cannot be identified with any known form of matter, which is why the adjective "exotic" is sometimes used to describe this presumed matter. However, in dismissing general relativity in favour of Newtonian gravitational theory for the study of galactic dynamics, insufficient attention has been paid to the fact that the stars that compose the galaxies are essentially in motion under gravity alone ("gravitationally bound"). It had been known for many years, in fact since the time of Eddington, that the gravitationally bound problem in general relativity is an intrinsically nonlinear problem even when the conditions are such that the field is weak and the motions are non-relativistic, at least in the time-dependent case. *Most significantly, we have found that under these conditions, the general relativistic analysis of the problem is also nonlinear for the stationary (non-time-dependent) case at hand.* Thus the intrinsically linear

Newtonian-based approach used to this point has been inadequate for the description of the galactic dynamics and Einstein's general relativity must be brought into the analysis within the framework of established gravitational theory This is an essential departure from conventional thinking on the subject and it leads to major consequences as we discuss in what follows. We will demonstrate that via general relativity, what we will refer to as "generating" potentials producing the observed flattened rotation curves can be linked to the mass density distributions of the essentially flattened disks, thus removing any necessity for dominant massive exotic dark matter halos in the total extended galactic composition. It should be stressed that this is insofar as accounting for the presently available data on rotation curves over the galactic near-central plane. However, further observations outside of this plane could possibly lead to different conclusions. We will discuss this as well in what follows.

There is another intriguing result that has emerged from our investigations. In our initial studies, we have found that in three galaxies, the threshold for luminosity, as we probe in the radial direction, occurs at about the same density for each: $10^{-21.75}$ kg·m^{-3}. Interestingly, we have now studied three additional galaxies and have found that as before, the optical luminosity edges occur at approximately the same density as prevailed in the first cases that we analyzed [11]. This concordance has the potential to further our understanding of galactic evolution.

Since our initial posting [1], many colleagues have offered their comments and criticisms. We address the most common areas of contention in Section 4. The totality of the issues known to us have been discussed in [2], [3], [4] and [5]. In those papers, we developed the theory in more detail and applied it to three galaxies and the Milky Way and in [11], to three additional galaxies. We also developed a new observational discriminator for assessing the degree, if any, of external matter that may lie beyond the visible/HI regions. This is determined by examining the rotation curves at different galactic latitudes, bringing into consideration the global dynamical structure of the galaxy. It is well to emphasize that the demand for global consistency applies not only to our own but also to all proposed models and theories.

AXIALLY-SYMMETRIC MODEL GALAXY

Using Newtonian theory, Mestel [12] considered a special rotating disk with surface density inversely proportional to radius. Using a disk potential with Bessel functions that we will also use in what follows but in quite a different manner, he found that it leads to an absolutely flat galactic rotation velocity curve. It also occurs for the MOND [7, 8, 9] model. It is particularly noteworthy that the gradient of the potential in this, as in all Newtonian treatments, relates to acceleration whereas in the general relativistic treatment, we will show that the gradient of a generating potential gives the stellar tangential velocity (14). This fact bears witness to the essentially different mathematical structure relating to the physics between the general relativity and the Newtonian gravity treatments of the free-fall problem.

When we consider the complexity of the detailed structure of a spiral galaxy with its arms and irregular density variations, it becomes clear that the modeling within the

context of the complicated theory of general relativity demands some simplifications. As long as the essence of the structure is captured, these simplifications are justified and we can derive valuable information. Thus, in terms of its essential characteristics, we consider a uniformly rotating fluid without pressure and symmetric about its axis of rotation. We do so within the context of general relativity. In generality, the stationary axially symmetric metric can be described in the form

$$ds^2 = -e^{\nu-w}(udz^2+dr^2) - r^2e^{-w}d\phi^2 + e^w(cdt - Nd\phi)^2 \tag{1}$$

where u, ν, w and N are functions of cylindrical polar coordinates r, z. It is easy to show that to the order required, u can be taken to be unity. Retaining terms of non-zero order in G for u induces terms of order G^n with $n > 1$ in the field equations. It is simplest to work in the frame that is comoving with the matter,

$$U^i = \delta^i_0 \tag{2}$$

where U^i is the four-velocity This is reminiscent of the standard approach that is followed for FRW cosmologies. However, the FRW spacetimes are homogeneous and they are not stationary. The comoving approach was taken in the pioneering paper by van Stockum [13] who set $w = 0$ from the outset. Interestingly, the geodesic equations imply that $w = constant$ (which can be taken to be zero as in [13]) even for the *exact* Einstein field equations as studied in [13]. In fact the requirement that $w = 0$ can be seen directly using (2) and the metric equation $g_{ik}U^iU^k = 1$ [2]. As in [14] [15], we perform a purely *local* (r,z held fixed at each point when taking differentials) transformation. It is to be noted that this local transformation is used only to deduce the connection between N and ω (and hence V). All subsequent work continues in the original unbarred comoving frame. The localized transformation is

$$\bar{\phi} = \phi + \omega(r,z)\,t \tag{3}$$

which locally diagonalizes the metric. In this way, we are able to determine the local angular velocity ω and the tangential velocity V as

$$\omega = \frac{Nce^w}{r^2e^{-w}-N^2e^w} \approx \frac{Nc}{r^2} \tag{4}$$

$$V = \omega r \tag{5}$$

with the approximate value applicable for the weak fields under consideration. The Einstein field equations to order G^1 with w kept for later comparison, are

$$\begin{aligned} 2r\nu_r + N_r^2 - N_z^2 &= 0, \\ r\nu_z + N_rN_z &= 0, \\ N_r^2 + N_z^2 + 2r^2(\nu_{rr}+\nu_{zz}) &= 0, \\ N_{rr} + N_{zz} - \frac{N_r}{r} &= 0, \end{aligned} \tag{6}$$

$$\left(w_{rr}+w_{zz}+\frac{w_r}{r}\right)+\frac{3}{4}r^{-2}(N_r^2+N_z^2)$$
$$+\frac{N}{r^2}\left(N_{rr}+N_{zz}-\frac{N_r}{r}\right)-\frac{1}{2}(\nu_{rr}+\nu_{zz}) \quad =8\pi G\rho/c^2 \tag{7}$$

where G is the gravitational constant and ρ is the mass density. Subscripts indicate partial differentiation with respect to the indicated variable. These equations are easily combined to give

$$\nabla^2 w+\frac{N_r^2+N_z^2}{r^2}=\frac{8\pi G\rho}{c^2} \tag{8}$$

where the first term is the flat-space Laplacian in cylindrical polar coordinates

$$\nabla^2 w\equiv w_{rr}+w_{zz}+\frac{w_r}{r} \tag{9}$$

and ν would be determined by simple integration.

The application of the freely gravitating constraint (i.e. stress-free motion) and the requirement that $w=0$ which arises from the choice of comoving coordinates, the field equations for N and ρ in this globally dust distribution are reduced to

$$N_{rr}+N_{zz}-\frac{N_r}{r}=0 \tag{10}$$

$$\frac{N_r^2+N_z^2}{r^2}=\frac{8\pi G\rho}{c^2}. \tag{11}$$

Note that with the minus sign in (10), N does *not* satisfy the Laplace equation. Note also that from both the field equation for ρ and the expression for ω that N is of order $G^{1/2}$. This is a point that has been misunderstood by some of our critics, leading them to erroneous conclusions.

The nonlinearity of the galactic dynamical problem is evident through the nonlinear relation between the functions ρ and N.. While we have eliminated w either by using the geodesic equations to get (11) or by the metric equation and the choice of comoving coordinates, N cannot be eliminated and hence nonlinearity is intrinsic to the study of the galactic dynamics. Rotation under freely gravitating motion is the key factor at play in the present problem. By contrast, for time-independence in the non-rotating problem, there must be pressure present to maintain a static configuration (thereby altering the right hand side of (6)), N vanishes for vanishing ω and $\nabla^2 w$ is non-zero yielding the familiar Poisson equation of Newtonian gravity. In the present case, it is the *rotation* as encapsulated through the function N that connects directly to the density. Thus the now nonlinear equation is in sharp contrast to the linear Poisson equation.

Interestingly, (10) can be expressed as

$$\nabla^2\Phi=0 \tag{12}$$

where

$$\Phi\equiv\int\frac{N}{r}dr \tag{13}$$

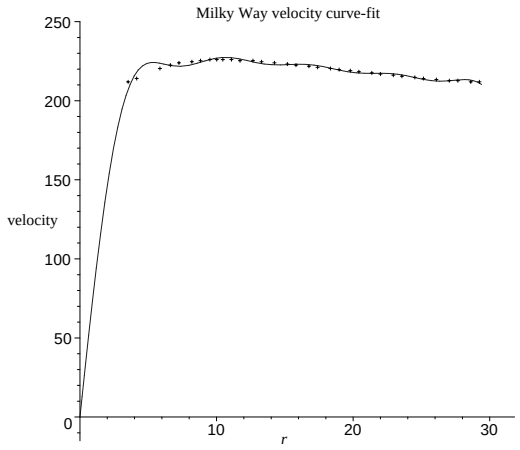


FIGURE 1. Velocity curve-fit for the Milky Way in units of m/s vs Kpc.

and hence flat-space harmonic functions Φ are the generators of the axially symmetric stationary pressure-free weak fields that we seek. (In fact Winicour [29] has shown that all such sources, even when the fields are strong, are generated by such flat-space harmonic functions.) These are the generating potentials referred to earlier. It is noteworthy that these generating potentials play a different role in general relativity than do the potentials of Newtonian gravitational theory even though both functions are harmonic. Using (5) and (13), we have the expression for the tangential velocity of the distribution

$$V = c\frac{N}{r} = c\frac{\partial\Phi}{\partial r}. \tag{14}$$

We now have the necessary mathematical framework in place.

CONNECTING THE OBSERVED GALACTIC ROTATION CURVES TO THE MODEL

We first consider the ideal strategy for galactic modelling, given the nature of the equations at hand. Since the field equation for ρ is nonlinear, the simpler way to proceed is to first find the required generating potential Φ and from this, derive an appropriate function N for the galaxy that is being analyzed. With N found, (11) yields the density distribution. If this agrees with the observations, the efficacy of the approach is established. This is in the reverse order of the standard approach to solving gravitational problems but it is most efficient in this formalism because of the existence of one linear field equation.

For any given galaxy, a suitable set of composing functions for the series that satisfies the linear equation is required. Once found, this yields the generating potential. With cylindrical polar coordinates, it is simplest to use separation of variables leading to the

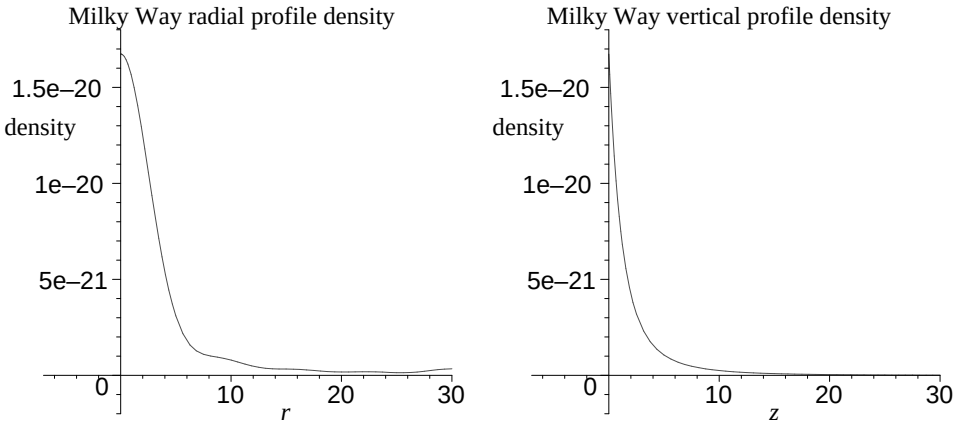


FIGURE 2. Derived density profiles in units of kg/m^3 for the Milky Way at (a) $z = 0$ and (b) $r = 0.001$ Kpc.

following solution for Φ in (12):

$$\Phi = Ce^{-k|z|}J_0(kr) \tag{15}$$

where J_0 is the Bessel function $m = 0$ of Bessel $J_m(kr)$ and C is an arbitrary constant (see for example [31]). Using this form of solution, the absolute value of z must be used to provide the proper reflection of the distribution for negative z. While this produces a discontinuity in N_z at $z = 0$, it is important to note that in the problem at hand, this discontinuity is consistent with the general case of having a density z-gradient discontinuity at the plane of reflection symmetry. This point has been the subject of considerable attention in the literature. We will return to this issue later.

The beauty of a linear equation is that it allows for linear superposition of solutions. From (12) we express the general solution of this form as the linear superposition

$$\Phi = \sum_n C_n e^{-k_n|z|}J_0(k_n r). \tag{16}$$

We choose n as required to the level of accuracy that we wish to achieve. From (16) and (14), the tangential velocity is

$$V = -c\sum_n k_n C_n e^{-k_n|z|}J_1(k_n r). \tag{17}$$

For this, we have used the Bessel relation $dJ_0(x)/dx = -J_1(x)$ (see, e.g. [32]). From (14), we see that if N should exceed r, the velocity would exceed c. This does not occur because with our choice of separable solutions, the velocity is given by (17). As r approaches 0, this function falls as r^1 (i.e. N approaches $r = 0$ as r^2) and so V falls to 0 as we see as well in the plots of the rotation curves. Thus the potential problem referred to by Wiltshire [30] is not present in our case. As well, for large r, the Bessel functions fall as $1/\sqrt{r}$ and hence the velocity goes to 0 for large r. In general, this would still lead to a large amount of matter external to the galaxy. However, by our selecting the right

multiplying coefficients in the Bessel solution sequence, we can achieve consistency with a very limited amount of external matter, far less than that which is indicated on the basis of Newtonian gravity. We have shown this by fitting fictitious faster-than-$1/\sqrt{r}$ fall-off data for the extension of the observed points in the velocity profile. We choose the k_n so that the $J_0(k_n r)$ terms are orthogonal to each other. The Bessel functions $J_0(kr)$ satisfy the orthogonality relation: $\int_0^1 J_0(k_n r)J_0(k_m r)rdr \mu \delta_{mn}$ where k_n are the zeros of J_0 at the limits of integration. With only 10 functions with parameters C_n, $n \in \{1 \ldots 10\}$, we have been able to achieve an excellent fit to the velocity curve for the Milky Way (see [4] for the curve-fitting coefficients).

A note should be added here regarding the set of constants C_n chosen for the rotation curve matching. These are *not* basic physical parameters as is the *single* physical parameter, the constant of universal gravitation G that links gravity with matter both in Newton's theory and Einstein's theory of gravity. Rather, these parameters are simply mathematical shaping numbers that adapt the solution of the Einstein equations to describe any given particular distribution of matter. Indeed the situation is analogous to the choosing of coefficients in a multipole expansion in classical physics that is designed to describe a distribution to any given level of accuracy.

The curve fit for the Milky Way is shown in Figure 1. It should be noted that the $J_1(x)$ Bessel functions are 0 at $x = 0$ and oscillate with decreasing amplitude, falling as $1/\sqrt{x}$ asymptotically [32]. However, this alone does not assure a realistic fall-off of matter. We deal with this issue in Section 4. Also, our curves drop as r approaches 0. This is in contrast to the Mestel [12] and MOND [7, 8, 9] curves that are flat everywhere. From (14) and (17), the N function is determined in detail and from (11), the density distribution follows. This is shown in Figure 2 as a function of r at $z = 0$ as well as a function of z at $r = 0.001$ Kpc. We see that the distribution is an essentially flattened disk with good correlation with the observed overall averaged density data for the Milky Way (see Figure 3). The integrated mass is found to be $21 \times 10^{10} M_\odot$ which is at the lower end of the estimated mass range of $20 \times 10^{10} M_\odot$ to $60 \times 10^{10} M_\odot$ as established by various researchers. It is to be noted that the approximation scheme would break down in the region of the galactic core should the core harbor a black hole or even a naked singularity (see e.g. [16]). *Most significantly, our correlation of the flat velocity curve is achieved with the modeled disk mass up to an order of magnitude smaller than the halo mass of exotic dark matter proposed by earlier studies.* (See e.g.[33, 34].

It is to be emphasized that general relativity does not distinguish between the luminous and non-luminous contributions. The deduced ρ density distribution is derived from the totality of the two. Any substantial amount of non-luminous matter (i.e. *conventional* non-exotic dark matter) would necessarily lie in the flattened region relatively close to $z = 0$ because this is the region of significant ρ and would be due to dead stars, planets, neutron stars and other normal non-luminous baryonic matter debris. Each term within the series has z-dependence of the form $e^{-k_n|z|}$ which causes the steep density fall-off profile as shown in Figure 2(b). This fortifies the picture of a standard galactic essentially flattened disk-like shape as opposed to a halo sphere. From the evidence provided thus far by rotation curves , there is no support for the widely accepted notion of the necessity for massive halos of exotic dark matter surrounding visible galactic disks: the generally accepted conventional gravitational theory, namely general relativity, can account for the

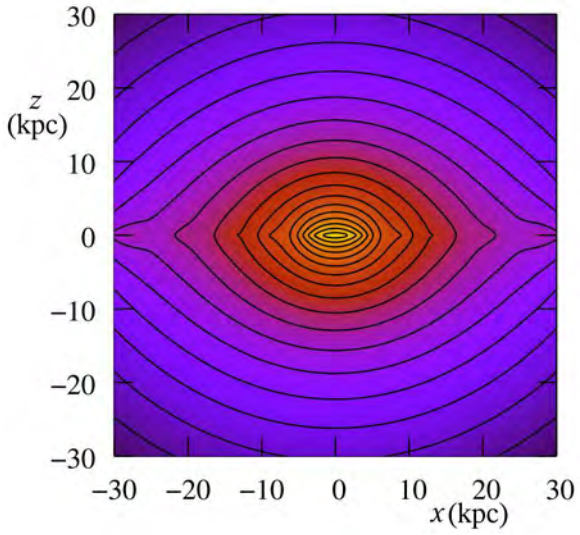


FIGURE 3. Cross-sectional density contour plot for the Milky Way model.

FIGURE 4. Velocity curve-fit and derived density for NGC 3031

observed flat galactic rotation curves linked to essentially flattened disks with no evident need for exotic dark matter, at least not with the velocity distribution data presently at our disposal.

We have also performed curve fits for the galaxies NGC 3031, NGC 3198 and NGC 7331. The data are provided in [4] and the remarkably precise velocity curve fits are shown in Figures 4 to 6 where the density profiles are presented for r at $z = 0$. Again the picture is consistent with the observations and the mass is found to be $10.1 \times 10^{10} M_\odot$ for NGC 3198. This can be compared to the result from Milgrom's [7, 8, 9] modified Newtonian dynamics of $4.9 \times 10^{10} M_\odot$ and the value given through observations (with Newtonian dynamics) by Kent [17] of $15.1 \times 10^{10} M_\odot$. While the visible light profile terminates at $r = 14$ Kpc, the HI profile extends to 30 Kpc. If the density is integrated to 14 Kpc, it yields a mass-to-light ratio of $7\Upsilon_\odot$. However, integrating through the HI outer region to $r = 30$ Kpc yields $14\Upsilon_\odot$ using data from [18].

For NGC 7331, we calculate a mass of $26.0 \times 10^{10} M_\odot$. Kent [17] finds a value of $43.3 \times 10^{10} M_\odot$. For NGC 3031, the mass is calculated to be $10.9 \times 10^{10} M_\odot$ as compared to Kent's value of $13.3 \times 10^{10} M_\odot$. Our masses are consistently lower than the masses projected by models invoking exotic dark matter halos and our distributions roughly tend to follow the contours of the optical disks. Most recently, we [11] have studied three additional galaxies and we have found the same trend of results.

In the course of our investigations, an interesting serendipitous discovery arose: From

FIGURE 5. Velocity curve-fit and derived density for NGC 3198

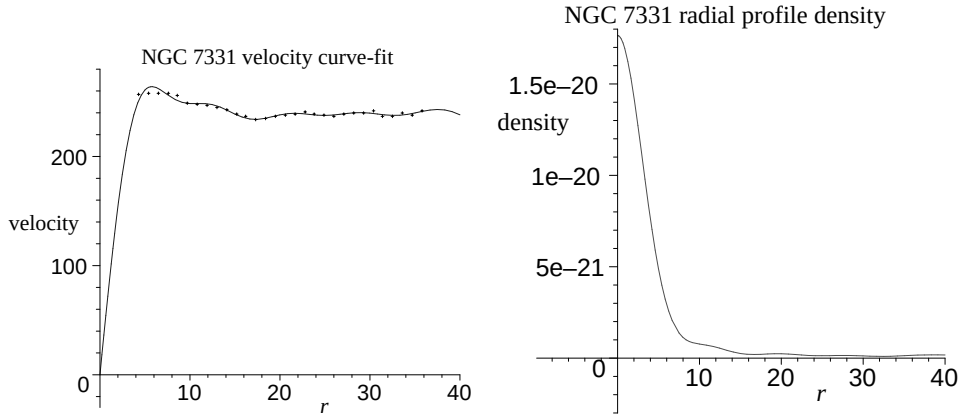


FIGURE 6. Velocity curve-fit and derived density for NGC 7331

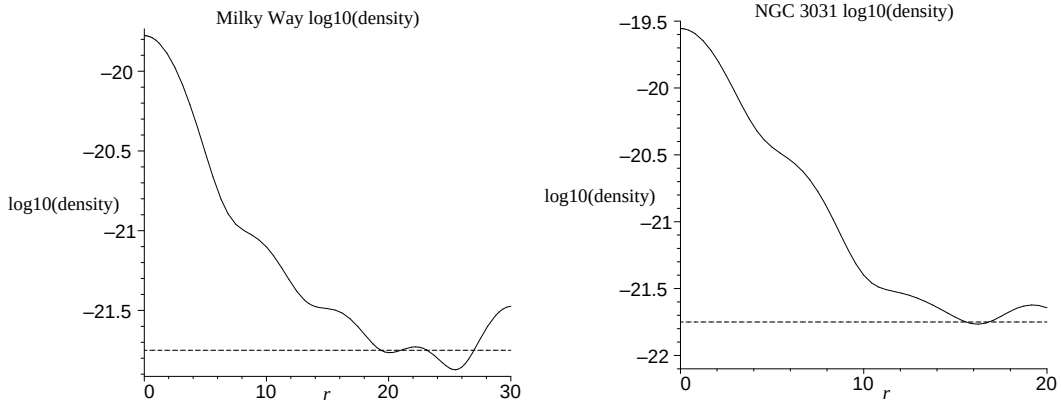


FIGURE 7. Log graphs of density for (a) the Milky Way and (b) NGC 3031 showing the density fall-off. The -21.75 dashed line provides a tool to predict the outer limits of visible matter. The fluctuations at the end are the result of limited curve-fitting terms.

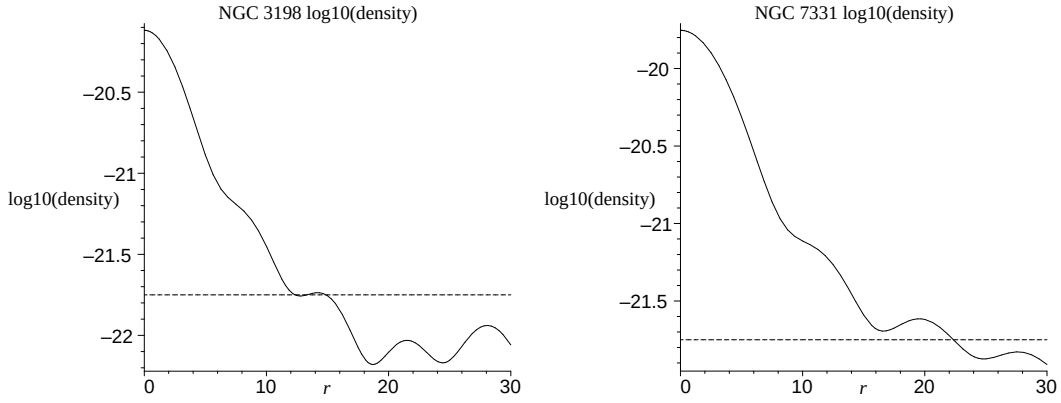


FIGURE 8. Log graphs of density for (a) the NGC 3198 and (b) NGC 7331 showing the density fall-off. The -21.75 dashed line provides a tool to predict the limits of luminous matter. As before, there are fluctuations near the border.

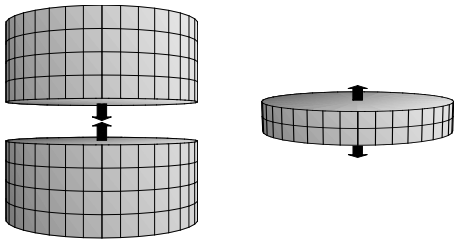


FIGURE 9. Normal vectors used to calculate flux

the figures provided by Kent [17] for optical intensity curves and our log density profiles for three galaxies NGC 3031, NGC 3198 and NGC 7331, we found that the threshold density for the onset of visible galactic light as we probed in the radial direction was the same for all, namely at approximately $10^{-21.75}$ kg·m^{-3} (Figure 7 and Figure 8). The concordance from our recent investigation [11] with three additional galaxies is moving in the direction of solidifying the indicated hypothesis that this density is the universal optical luminosity threshold for galaxies as tracked in the radial direction. Alternatively, should this hypothesis be further substantiated, the radius at which the optical luminosity fall-off occurs can be predicted for other sources using this special density parameter. The predicted optical luminosity fall-off for the Milky Way is at a radius of 19-21 Kpc based upon the density threshold indicator that we have determined.

CRITICAL CHALLENGES AND OUR REPLIES

There has been an interesting variety of papers critical of our work as well as two papers lending support to our work [19] [20]. Some of the issues arising from these papers have been addressed in [2], in [3] and in [4]. In this paper, we will focus on the key areas of criticism. The complete analysis is contained in our earlier papers and in the book [5].

An issue first raised privately to us by some colleagues and later in [21] [22] concerns the nature of the matter distribution. They have noted that given the existence of the discontinuity of N_z that we had pointed to in [1], a significant surface tensor S_i^k can be constructed with a surface density component given by

$$(8\pi G/c^2)S_t^t = \frac{N[N_z]}{2r^2} - \frac{[\nu_z]}{2} \tag{18}$$

to order G^1. The notation $[..]$ denotes the jump over a discontinuity of the given function, here at $z = 0$. Using (6), this becomes

$$(8\pi G/c^2)S_t^t = \frac{N[N_z]}{2r^2} + \frac{N_r[N_z]}{2r} \tag{19}$$

It was claimed that this necessarily implied the existence of a singular *physical* surface of mass in the galactic plane above and beyond the continuous mass distribution that we had found, thus rendering our model unphysical.

This challenge prompted us to calculate the surface mass that was said to be present in the first four galaxies that we had studied. This was done by integrating (19) over

the surface. We first did so without paying attention to the actual sign of the result. Suspicions were aroused from the discovery that (19) in each case gave a numerical value slightly less than the mass that we had derived from the volume integral of our *continuous* mass density distribution using (17), (14) and (11). This pointed to a plausible explanation: in our case, *with our choice of model*, there is no *physical* mass layer present on the $z = 0$ plane. *The surface integral of this singular layer is merely a mathematical construct that indirectly describes most of the continuously distributed mass by means of the Gauss divergence theorem*. To see this, consider the vector $\mathbf{F}$ defined as

$$\mathbf{F} \equiv A(r,z)\mathbf{e}_r + B(r,z)\mathbf{e}_z \tag{20}$$

where

$$(8\pi G/c^2)B \equiv \frac{NN_z}{2r^2} + \frac{N_r N_z}{2r} \tag{21}$$

as a first option. We choose $A(r,z)$ so that

$$\int \nabla \cdot \mathbf{F} dV \equiv (8\pi G/c^2)M \tag{22}$$

where M is the total mass. As a more transparent second option, we choose

$$(8\pi G/c^2)B \equiv \frac{NN_z}{r^2} \tag{23}$$

where we define

$$\nabla \cdot \mathbf{F} \equiv (8\pi G/c^2)\rho \tag{24}$$

From these definitions, we deduce the form of $A(r,z)$ in order to produce the density as expressed through N in (11). We calculate the mass over the cylindrical volume defined by $-\infty < z < \infty$, $0 < r < r_{galaxy}$. By the Gauss divergence theorem, the volume integral of ρ, via (24) is equal to the integral of the normal component of $\mathbf{F}$ over the bounding surfaces. However, for the application of the Gauss theorem, the integration must be over a continuous domain and since the $\mathbf{e}_z$ component is discontinuous over the $z = 0$ plane, the volume integral must be split into an upper and a lower half. The two new surface integrals together would constitute the jump integral of (19) in the first option if one were to be cavalier about the directions of unit *outward* normals, as we shall discuss in what follows. The surfaces above and below the galaxy give zero because of the exponential factors in z and the final small contribution comes from the cylinder wall via the A function.

In our solution, the actual *physical* distribution of mass is not in concentrated layers over bounding surfaces: the Gauss theorem gives the value of the *distributed* mass via equivalent purely mathematical surface constructs as we are familiar from elementary applications of this theorem. Physically, the density is well defined and continuous throughout, except on the $z = 0$ plane. In fact the limits as $z = 0$ is approached give the same finite values from above and below. While the field equations break down at $z = 0$, the density for a physically viable model is logically defined by this limit at $z = 0$. However, with the chosen form of solution, the density *gradient* in the z direction is discontinuous on the $z = 0$ plane. This gradient undergoes a reversal for a galactic

distribution with diminishing density in both directions away from the symmetry plane. It is most convenient to achieve this with an abrupt reversal as we have done. There is no indication that this choice alters the essential physics.

Thus we have shown via the Gauss divergence theorem, that the supposed surface layer is merely a re-expression of the integrals that constitute the *continuous volume distribution* of mass. Indeed if one were to reject this interpretation and insist that these surface integrals reveal additional mass in the form of a layer, then the Gauss theorem would indicate that this mass must be negative. Indeed various authors (e.g. [22] [24]) have referred to negative mass layers. However, as Bondi had emphasized in his writings, negative mass repels rather than attracts. Therefore we had set out to test the viability of the presence of such negative mass to see if repulsion rather than attraction was in evidence. We considered a test particle in our model that was comoving with the rotating dust apart from having a component of velocity U^z normal to the $z = 0$ plane. The geodesic equation in the z direction reduces to

$$\frac{dU^z}{ds} = \frac{N_r N_z (U^z)^2}{2r}. \tag{25}$$

We had computed the complete N series for the galaxy NGC7331 (see [1]). We then focused upon points in the range $r = 0.1$ to 30 and points above the $z = 0$ symmetry plane $z = 0.001$ to 1 for the right hand side of (25). All of the points gave a negative value as expected for the z acceleration (i.e. attraction) of a particle in the region above the symmetry plane. However, if the $z = 0$ surface actually harboured a *physical* negative mass surface layer, indeed one of numerical value comparable to the positive mass of the normal galactic distribution, then at the very least, one would have expected to witness a *repulsion* of the particle as the test particle approached the boundary, indeed given the extent of the supposed negative mass, a violent repulsion. The absence of this occurrence adds further support to our original model [1] as being free of surface layers of mass.

It is true that our choice of solution leads to a discontinuity in the z-derivative of N across the $z = 0$ plane. It is well to reiterate and emphasize the argument: it goes hand-in-hand with the physically natural density *gradient* discontinuity across the symmetry plane. To see this in another way, consider the essential characteristics of our model which consists of dust with reflection symmetry about the $z = 0$ plane. The density naturally increases symmetrically as this plane is approached from above and from below with the same absolute value but opposite sign from symmetry. In all generality, the density z-gradient will be different from zero as this plane is approached and because of reflection symmetry, this gradient will of necessity be discontinuous.

It is to be noted that the density gradient is governed by the behavior of odd derivatives of N with respect to z. However, the density itself is governed by N_z^2 (11) which has the same limit as z approaches 0 from above or below. Thus, we define the value of ρ at $z = 0$ by this common limit and hence the singularity is removable. It is only with delicate fine-tuning that this discontinuity can be avoided and this will be the case only if the density gradient is adjusted to be precisely zero as the $z = 0$ plane is approached from above and below.

As an exercise in response to critical comments [2], we achieved this approximately by choosing $\cosh(\kappa_n z)$ functions in place of exponential functions to span the region in a sandwich encompassing the symmetry plane and employing the usual exponential

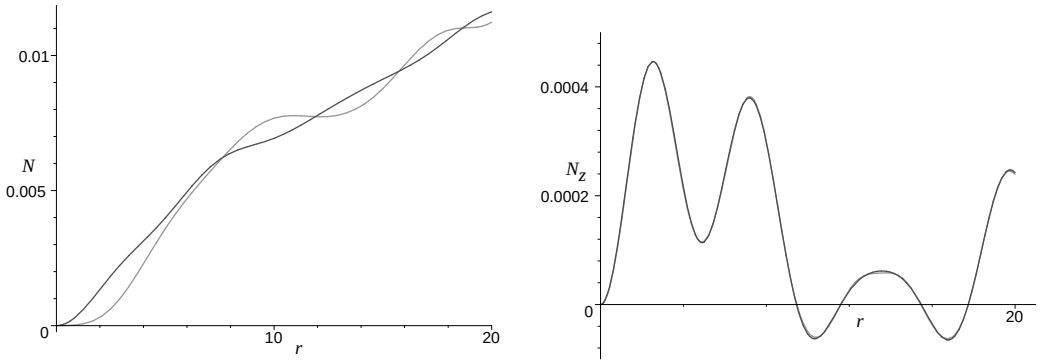


FIGURE 10. Matching conditions for N and $\partial N/\partial z$ at $z = z_0$.

functions beyond this sandwich. This led to the issue of matching the N and N_z functions along the external/internal region joins and it was achieved by using many different k_n parameters for the external exponential functions as opposed to the original 10 internal parameters of the original model. In [22], it was claimed that a matching could not be achieved but these authors had not realized that we used different and many parameters for the outside regions. Since then, we have refined the fit further by employing hundreds of external parameters and the improved fit is shown in Figure 10. However, it must be stressed that the generic situation would be one in which the density gradient is discontinuous at $z = 0$.

In a follow-up paper [23], the authors pasted a finite thickness layer of density and stress as a sandwich about the $z = 0$ symmetry plane. This is of some interest in building more general galactic models. However, there remains the assertion that when the sandwich is reduced to zero thickness, a surface layer arises which, by the right choice of parameter, results in having the layer consist of negative mass. Again, the fact that test particles are attracted towards rather than repelled from this layer, regardless of the assumed sign of the parameter, negates this interpretation. An essential point is this: singularities can arise in many forms and must be interpreted properly with the physics in mind.

In [25], the well-known expression of the field equations in the harmonic gauge in Cartesian coordinates

$$\partial_k \partial^k h^{ab} = \frac{16\pi G}{c^4} \tau^{ab} \tag{26}$$

(τ^{ab} includes the energy-momentum tensor of the matter plus the nonlinear terms in the Einstein equations) is invoked. (A related line of reasoning was followed in [21]). In [25], the author presents the standard description of the post-Newtonian perturbation scheme to conclude that the solution to the galactic problem must be the usual Newtonian one and that all corrections must be of higher order. Firstly, we did not use this scheme (as noted as well in [26]). Just as one would not logically choose Cartesian coordinates in the harmonic gauge to describe FRW cosmologies, one would not normally choose these for our stationary axially symmetric galactic problem. Our problem is greatly simplified with cylindrical polar coordinates comoving with the matter. Secondly, for the gravitationally bound system under study, the metric components are of

different orders in G. This is a key point that was overlooked by some of our critics. If one were to take the approach suggested in [25], the equations (26) could be schematically expressed as

$$\nabla^2 h_{(1/2)} = 0, \quad \nabla^2 h_{(1)} = GT + h^2_{(1/2)} \tag{27}$$

where tensorial superscripts have been suppressed and the lower case numbers refer to orders in G. In this manner, we would have incorporated the nonlinear structure of our system within the framework of the scheme suggested by [25]. The novel aspect is that the lowest order equation (of order $G^{1/2}$) in (27) has zero on the RHS and the second equation that would normally be the Newtonian Poisson equation, differs in that it has nonlinear terms. Thus, the structure of our solution does not proceed as in the standard approach of (26). In the latter standard approach, the lowest order base solution is the Newtonian solution whereas in the galactic problem, the lowest order equation is the Laplace equation for which an order $G^{1/2}$ solution is necessary (see [27] where this component is inappropriately chosen to be zero) and the next order (order G^1) equation for the density (28),

$$\frac{N_r^2 + N_z^2}{r^2} = \frac{8\pi G\rho}{c^2} \tag{28}$$

has nonlinear terms in the metric in the form of the squares of the derivatives of an order $G^{1/2}$ metric tensor component N. Thus, our situation is unlike standard iterative perturbation scheme applications as envisaged in [25]. Hence there is no basis to draw the conclusions that are expressed therein.

Further in [25], the author refers to "extra matter" in the symmetry plane of the galaxy and muses whether our model "could be somehow fixed". However, in [2] we presented the evidence that our solution embodies the physically natural density gradient discontinuity at the plane of symmetry and that it does not contain extra matter. Moreover, we showed that if there were to be a surface layer of mass, it would be negative mass but this was negated by the attraction rather than repulsion of test particles near the symmetry plane, as we discussed above.

With regard to the issue of gauge, it was argued in [21] that asymptotically flat solutions are unattainable with a lead-off $G^{1/2}$ order metric component. However, we have shown that they are readily attainable in conjunction with the physically desirable N_z discontinuity and are approximately attainable with the smoothed fine-tuned solution discussed above. Moreover, they are precisely attainable when an essential singularity is invoked. This was almost achieved in [19]. Their axis singularity prevented global asymptotic flatness. However, exact solutions with compactified singularities of the Weyl type are likely to rectify this deficiency.

A key point is that the equations have an inherent nonlinearity as a result of the fact that the metric components are of different orders and the different orders are a necessary consequence of the problem being a gravitationally bound one. The authors of [19] arrive at our equations (6), (7) (with w set to zero) apart from the exponential ν factor which they later note can be taken to be a constant scaling factor and find the same order of magnitude reduction of galactic mass that we had found [1] starting from their exact solution class. This provides some vindication for our analysis. It should be noted that their scaling factor is actually incorporated in our solutions within the computed

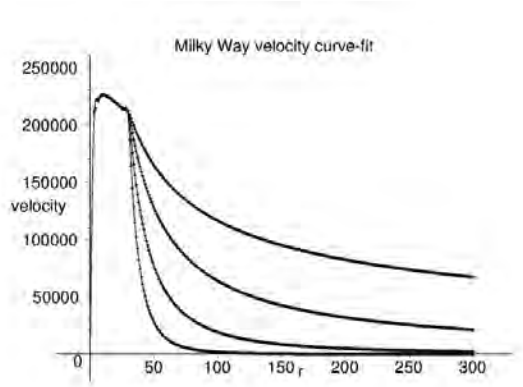


FIGURE 11. Beyond the HI region, the velocity can be modeled in many different manners: here $V \mu 1/\sqrt{r}$, $V \mu 1/r$, $V \mu 1/r^2$ and $V \mu 1/r^4$ are illustrated.

amplitudes of our basis expansion functions. To be particularly noted in [19] is that their solution class is fine-tuned as the density gradient is precisely zero at $z = 0$. The price that is paid to achieve this degree of smoothness is the incorporation of an axial singularity. These authors justify the singularity by identifying it as a jet. While jets are observed in various galaxies in their formative stages, they are not known to be present in the essentially stationary galaxies that are being modeled with this class.

We display the accumulated mass for the Milky Way in a highly extended cylindrical volume of 300 Kpc in size in figure 12. It is to be noted that even assuming a Newtonian-like fall-off of the form $1/\sqrt{r}$, there is a far less amount of accumulated mass up to a radius of ten times the visible radius than is envisaged by the use of Newtonian as opposed to general relativistic galactic dynamics. An even slower accumulation of mass is seen for the $1/r$ fall-off. For such a fall-off, the accumulated mass is approximately $35 \times 10^{10} M_{\odot}$ at a radius of 300 Kpc and a linear extrapolation to r=900 Kpc yields a value of $39.2 \times 10^{10} M_{\odot}$, a very modest increase in comparison to Newtonian modeling. Moreover, the faster fall-offs of $1/r^2$ and $1/r^4$ yield very minor mass increases out to very large radii as can be seen in figure 12. Note from this figure that the accumulated mass at 30 Kpc is approximately the same for the various fall-off scenarios as well as the value stated in Section 3 where we used only 10 parameters and where we did not focus on the bahaviour of the model beyond the 30 Kpc edge of the HI region.

This fortifies our contention that general relativity obviates the need for overwhelmingly dominant massive halos of exotic dark matter.

It is to be noted that our models are *globally* dust. (We make this choice for the composition and distribution of the matter for the purpose of mathematical simplicity.) Therefore there is no basis for a matching with the vacuum Kerr metric asymptotically. Our models are asymptotically flat with a well-defined mass. The Tolman integral dictates the value of this mass and since there is no stress and the fields are weak, the Tolman mass to lowest order is simply given by the coordinate volume integral of the density.

In [28], the authors fault our models as extended constructs that indicate enormous quantities of mass beyond the HI regions. This is a useful point of criticism in that we

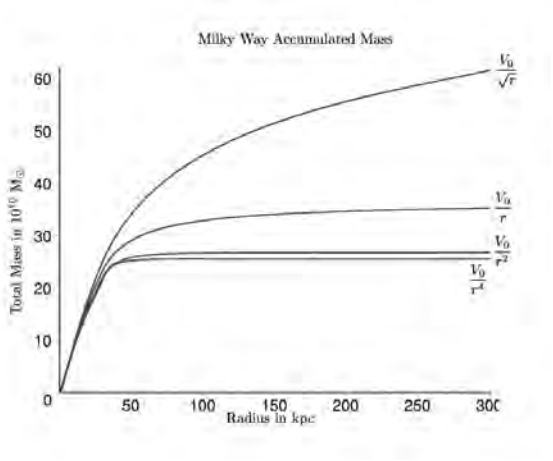


FIGURE 12. The Milky Way's accumulated mass as a consequence of velocity fall-off beyond the HI region.

had not investigated earlier the asymptotic consequences of the particular parameter sets that we had chosen to model the observed rotation curves. It is to be noted however that their argument that mass accumulates linearly in r is faulty as a generalization. With the correct combination of parameters, the term that would lead to such an accumulation can be eliminated. Our examples in which we achieve minimal accumulation, provide the direct proof that this is the case. However, as we first reported in [3], we assure more realistic fall-off scenarios. It is to be noted that in so doing, while the expansion parameters are no longer the same as in the earlier sets, we have determined that the net physical effects are of insignificant difference within the observed matter distribution in the two approaches. We find that the accumulated mass profiles indicate that most of the mass of a galaxy is confined fairly close to the region of the visible disk with *modest* accumulations of mass beyond this region. General relativity achieves this with a pressure-free fluid model, unlike Newtonian gravity. In an interesting approach from a very different direction, Lusanna [20] has pointed to relativistic inertial effects that do not have a Newtonian limit counterpart. He has suggested that in the weak field limit, these effects could match our results.

A VELOCITY DISPERSION TEST AS A DISCRIMINANT FOR EXTRA MATTER

Clearly it is important to approach the exotic dark matter issue in as many ways as possible. After all, from a purely formal point of view, general relativity should be able to model vastly extended distributions of pressure-free fluids in rotation. In this vein, we have constructed a test in principle that relies upon data in the *visible/HI* regime, thus making it particularly useful. When we examine Figure 12, we see that different constructed velocity fall-off profiles beyond the HI region imply different mass accumulations in those external regions. Carrying these back with continuity into the

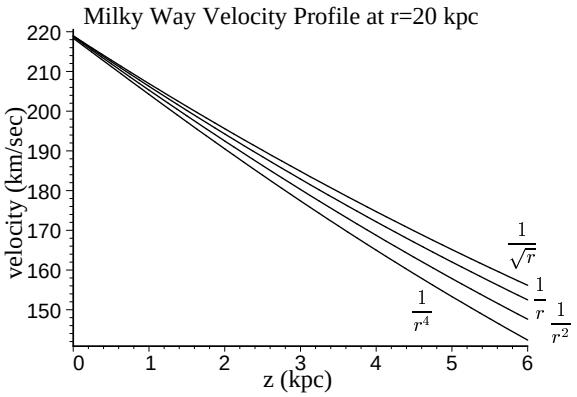


FIGURE 13. Velocity dispersion at $r = 20$ kpc for the Milky Way.

visible/HI region, we find that the extent of the velocity dispersion as we track curves at different non-zero z values depends on the assumed external velocity profile fall-off. (See, for example. Figure 13.)

With sufficient data, it should be possible, at least in principle, to provide limits on the extent of extra matter that might lie outside of the visible/HI region. To this point, we have only the data provided in [35] [36] [37] but far more data will be required to provide an adequate discriminating test.

CONCLUDING COMMENTS

We are often challenged by interested readers to justify how our results could be so different from Newtonian predictions. They note that the dynamic solar system analysis proceeds very accurately on the basis that the planets move in almost the same manner as deduced by general relativity as that deduced by Newtonian dynamics. The observations of their motions confirm this. However there is an essential difference between the solar system dynamics and the galactic dynamics. In the case of the solar system, the primary source of gravity is the sun and the planets are treated as *test* particles in this field apart from contributing minor perturbations when the slight changes are being sought. The planets respond to the field of the sun but their own gravitational contributions are not retained since they are so small. By contrast, in the galaxy problem, the source of the field is the combined rotating mass of all of the freely-gravitating elements themselves that compose the galaxy. There is no one single dominant contributor in the galactic problem.

We have seen that the nonlinearity for the computation of density distribution inherent in the Einstein field equations for a stationary axially-symmetric pressure-free mass distribution, even in the case of weak fields, leads to the correct galactic velocity curves as opposed to the incorrect curves that had been derived on the basis of Newtonian gravitational theory. Indeed the results were consistent with the observations of velocity as a function of radius plotted as a rise followed by an essentially flat extended region and no large massive halos were required to achieve them. The density distribution

that is revealed thereby is one of an essentially flattened disk without an accompanying massive vastly extended dark matter halo. With the "dark" matter being associated with the disk which is itself visible, it is natural to regard the non-luminous material as normal baryonic matter. In a very recent investigation, Auping [38] has tallied the baryonic mass contributions from the galactic elements and has found consistency with our results.

To some extent we have had to assume extensions of matter distribution with assumptions regarding ultimate velocity fall-offs beyond that which is actually observed in order to make comparisons. In the course of these investigations, we have seen that these can readily yield a picture of galactic structure without huge extended massive halos of exotic dark matter. It would be helpful if new data beyond those presently available would be produced. This would help tie down the complete physical picture.

Of particular interest is that we have within our grasp a criterion for determining the extent, if of any significance, of extra matter beyond the visible and HI regions of a galaxy. It is possible in principle to determine this with data solely *within* the visible/HI region by plotting the velocity dispersion of rotation curves for various z values. This is an attractive area for future research. In particular, it expands the demands upon not only our galactic model but also upon any other proposed model by other researchers. It asks for consistency between observation and theoretical prediction for the overall averaged picture of stellar motions within the complete galaxy.

Nature is merciful in providing one linear equation that enables us by superposition to model disks of variable density distributions. This opens the way to studies of other sources and with further refinements. It is to be emphasized that what we have taken is a first step, a general relativistic as opposed to a Newtonian analysis at the galactic scale. It is noteworthy that others have now come to recognize that the galactic problem is a nonlinear general relativistic problem even given the conditions of weak fields and non-relativistic velocities. It will be of interest to extend this general relativistic approach, with the hitherto neglected consideration of nonlinearities, to the other relevant areas of astrophysics with the aim of determining whether there is any scope remaining for the presence of any exotic dark matter in the universe. We have taken one further step in this direction by analyzing an *intrinsically dynamic* free-fall model, that of an idealized Coma Cluster of galaxies [39] (see accompanying article). This work has the following attractions: firstly it makes use of an *exact* solution of the Einstein equations, hence removing any issues arising from the application of approximations and secondly, being free of any singularities of any kind, it eliminates any possibility of the kind of objections that we have discussed in the present paper.

We share the belief of many that the scientific method has been most successful when guided by Occam's razor, that new elements should not be introduced into a theory unless absolutely necessary. If dark matter should turn out to be another case similar to the ether of the 19th Century, it is well for us to determine this sooner rather than later.

REFERENCES

1. F. I. Cooperstock and S. Tieu, astro-ph/0507619.
2. F. I. Cooperstock and S. Tieu, astro-ph/0512048.
3. F. I. Cooperstock and S. Tieu, *Mod. Phys. Lett. A.* **21**, 2133 (2006).

4. F. I. Cooperstock and S. Tieu, *Int. J. Mod. Phys. A.* **22**, 2293 (2007).
5. F. I. Cooperstock, *General Relativistic Dynamics: Extending Einstein's legacy throughout the universe* (World Scientific, Singapore, 2009).
6. J. Binney and S. Tremaine, *Galactic Dynamics* (Princeton University Press, Princeton, N.J., 1987).
7. M. Milgrom, *Ap. J.* **270**, 365 (1983).
8. M. Milgrom and J. D. Bekenstein, *Ap. J.* **286**, 7 (1984).
9. J. D. Bekenstein, *Developments in General Relativity, Astrophysics and Quantum Theory* (Eds. F. I. Cooperstock, L. P. Horwitz and J. Rosen, IOP Publishing Ltd, Bristol, England, 1990).
10. J. R. Brownstein and J. W. Moffat, astro-ph/0506370.
11. J. Carrick and F. I. Cooperstock, astro-ph/1101.3224.
12. L. Mestel, *MNRAS* **126**, 553 (1963).
13. W. J. van Stockum, *Proc. R. Soc. Edin.* **57**, 135 (1937).
14. J. M. Bardeen, *Astrophys. J.* **162**, 71 (19700.
15. W. B. Bonnor, *J. Phys. A: Math. Gen.* **10**, 1673 (1977).
16. F. I. Cooperstock, S. Jhingan, P. S. Joshi and T. P. Singh, *Class. Quantum Grav.* **14**, 2195 (1997).
17. S. M. Kent, *Astron. J.* **93**, 816 (1987).
18. T. S. van Albada, J. N. Bahcall, K. Begeman and R. Sanscisi, *Ap. J.* **295**, 305 (1985).
19. H. Balasin and D. Grumiller, astro-ph/0602519.
20. L. Lusanna, gr-qc/0604120.
21. M. Korzynski, astro-ph/0508377.
22. D. Vogt and P. S. Letelier, astro-ph/0510750.
23. D. Vogt and P. S. Letelier, astro-ph/0611428.
24. W. B. Bonnor, 2005 preprint.
25. D. Garfinkle, *Class. Quant. Grav.* **23**, 1391 (2006), gr-qc/0511082.
26. M. D. Maia, A. J. S. Capistrano and D. Muller, astro-ph/0605688.
27. D. J. Cross, astro-ph/0601191.
28. D. Menzies and G. J. Mathews, gr-qc/0604092.
29. J. Winicour, *J. Math. Phys.* **16**, 1805 (1975).
30. D. L. Wiltshire, gr-qc/0702082.
31. J. C. N. de Araujo and A. Wang, *Gen. Rel. Grav.* **32**, 1971 (2000).
32. L. B. Ford, *Differential Equations* (McGraw-Hill, 1955).
33. L. Clewley, S. J. Warren, P. C. Hewett, M. Wilkinson and W. N. Evans, astro-ph/0310675.
34. M. Wilkinson and N. Evans, *MNRAS* **310**, 645 (1999).
35. F. Fraternali et al, astro-ph/0410375.
36. G. Battaglia et al, *MNRAS* **364**, 433 (2005).
37. G. Heald, PhD Thesis, University of New Mexico, 2006.
38. J. Auping, private communication.
39. F. I. Cooperstock and S. Tieu, *Mod. Phys. Lett. A.* **23**, 1745 (2008).

Modified Gravity Or Dark Matter?

J. W. Moffat

Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada
Department of Physics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

Abstract. Modified Gravity (MOG) has been used successfully to explain the rotation curves of galaxies, the motion of galaxy clusters, the Bullet Cluster, and cosmological observations without the use of dark matter or Einstein's cosmological constant. We review the main theoretical ideas and applications of the theory to astrophysical and cosmological data.

Keywords: Cosmology, modified gravity, dark matter
PACS: 04.20.Cv,04.50.Kd,04.80.Cc,98.80.-k

1. INTRODUCTION

The ingredients of the standard model of astrophysics and cosmology are:

1. General Relativity,
2. Large-scale homogeneity and isotropy,
3. 5% ordinary matter (baryons and electrons),
4. 25% dark matter,
5. 70% dark energy,
6. Uniform CMB radiation, $T \sim 2.73$ degrees,
7. Scale-free adiabatic fluctuations $\Delta T/T \sim 10^5$.

Although the model fits available astrophysical and cosmological data [1], it opens up the mystery that about 95% of all matter and energy are invisible. The dark matter (DM) does not interact with ordinary baryonic matter and light. No current experiments have succeeded in detecting DM . The SN1a supernovae data [2, 3] have created the need for the expansion of the universe to accelerate, promoting the mechanism of dark energy.

In the event that DM is not detected, then to fit all available astrophysical and cosmological data, we are required to modify Newtonian and Einstein gravity without assuming the undetected DM. A modified gravity (MOG) theory, also known as Scalar-Tensor-Vector Gravity or STVG [4, 5, 6], is based on an action that incorporates, in addition to the Einstein-Hilbert term and the matter action, a massive vector field, and three scalar fields corresponding to running values of the gravitational constant, the vector field coupling constant, and the vector field mass.

We begin in Section 2 by introducing the theory through the action principle, and establish key assumptions that allow us to analyze physically relevant scenarios. In Section 3, we derive the field equations using the variational principle. In Section 4 we solve the field equations in the static, spherically symmetric case. In Section 5, we postulate the action for a test particle, and obtain approximate solutions to the field equations for

a spherically symmetric gravitational field. In Section 6 we demonstrate how the Friedmann equations of cosmology can be obtained from the theory. In Section 7, we utilize the theory to obtain estimates for galaxy rotation curves, galaxy cluster dynamics and show how the solutions we obtained for the field equations remain valid from cosmological to solar system scales. Lastly, we end in Section 8 with conclusions.

2. MODIFIED GRAVITY THEORY

The action of our theory is constructed as follows [4]. We start with the Einstein-Hilbert Lagrangian density that describes the geometry of spacetime:

$$\mathscr{L}_G = -\frac{1}{16\pi G}(R+2\Lambda)\sqrt{-g}, \tag{1}$$

where G is the gravitational constant, g is the determinant of the metric tensor $g_{\mu\nu}$ (we are using the metric signature $(+,-,-,-)$), and Λ is the cosmological constant. We set the speed of light, $c=1$. The Ricci-tensor is defined as

$$R_{\mu\nu} = \partial_\alpha \Gamma^\alpha_{\mu\nu} - \partial_\nu \Gamma^\alpha_{\mu\alpha} + \Gamma^\alpha_{\mu\nu}\Gamma^\beta_{\alpha\beta} - \Gamma^\alpha_{\mu\beta}\Gamma^\beta_{\alpha\nu}, \tag{2}$$

where $\Gamma^\alpha_{\mu\nu}$ is the Christoffel-symbol, while $R = g^{\mu\nu}R_{\mu\nu}$.

We introduce a "fifth force" vector field ϕ_μ via the Maxwell-Proca Lagrangian density:

$$\mathscr{L}_\phi = -\frac{1}{4\pi}\omega\left[\frac{1}{4}B^{\mu\nu}B_{\mu\nu} - \frac{1}{2}\mu^2\phi_\mu\phi^\mu + V_\phi(\phi)\right]\sqrt{-g}, \tag{3}$$

where $B_{\mu\nu} = \partial_\mu\phi_\nu - \partial_\nu\phi_\mu$, μ is the mass of the vector field, ω characterizes the strength of the coupling between the "fifth force" and matter, and V_ϕ is a self-interaction potential.

Next, we promote the three constants of the theory, G, μ,ω, to scalar fields by introducing associated kinetic and potential terms in the Lagrangian density:

$$\begin{aligned}\mathscr{L}_S &= -\frac{1}{G}\left[\frac{1}{2}g^{\mu\nu}\left(\frac{\nabla_\mu G\nabla_\nu G}{G^2} + \frac{\nabla_\mu\mu\nabla_\nu\mu}{\mu^2} - \nabla_\mu\omega\nabla_\nu\omega\right)\right.\\ &\left. + \frac{V_G(G)}{G^2} + \frac{V_\mu(\mu)}{\mu^2} + V_\omega(\omega)\right]\sqrt{-g},\end{aligned} \tag{4}$$

where ∇_μ denotes covariant differentiation with respect to the metric $g_{\mu\nu}$, while V_G, V_μ, and V_ω are the self-interaction potentials associated with the scalar fields.

Our action integral takes the form

$$S = \int (\mathscr{L}_G + \mathscr{L}_\phi + \mathscr{L}_S + \mathscr{L}_M)\, d^4x, \tag{5}$$

where $\mathscr{L}_M$ is the ordinary matter Lagrangian density, such that the energy-momentum tensor of matter takes the form:

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}}\frac{\delta S_M}{\delta g^{\mu\nu}}, \tag{6}$$

where $S_M = \int \mathscr{L}_M \, d^4x$. A "fifth force" matter current can be defined as:

$$J^\nu = -\frac{1}{\sqrt{-g}} \frac{\delta S_M}{\delta \phi_\nu}. \tag{7}$$

We assume that the variation of the matter action with respect to the scalar fields vanishes:

$$\frac{\delta S_M}{\delta X} = 0, \tag{8}$$

where $X = G, \mu, \omega$.

3. FIELD EQUATIONS

The field equations of the theory can be obtained in the form of the first and second-order Euler-Lagrange equations [5]:

$$\frac{1}{4\pi}\left[\omega \nabla_\mu B^{\mu\nu} + \nabla_\mu \omega B^{\mu\nu} + \omega \mu^2 \phi^\nu - \omega \frac{\partial V_\phi(\phi)}{\partial \phi_\nu}\right] = J^\nu, \tag{9}$$

$$\nabla^\nu \nabla_\nu \mu - \frac{\nabla^\nu \mu \nabla_\nu \mu}{\mu} - \frac{\nabla^\nu G \nabla_\nu \mu}{G} + \frac{1}{4\pi} G \omega \mu^3 \phi_\mu \phi^\mu + \frac{2}{\mu} V_\mu(\mu) - V'_\mu(\mu) = 0, \tag{10}$$

$$\nabla^\nu \nabla_\nu \omega - \frac{\nabla^\nu G \nabla_\nu \omega}{G} - \frac{1}{8\pi} G \mu^2 \phi_\mu \phi^\mu + \frac{G}{16\pi} B^{\mu\nu} B_{\mu\nu} + \frac{1}{4\pi} G V_\phi(\phi) + V'_\omega(\omega) = 0, \tag{11}$$

$$\nabla^\nu \nabla_\nu G - \frac{3}{2}\frac{\nabla^\nu G \nabla_\nu G}{G} + \frac{G}{2}\left(\frac{\nabla^\nu \mu \nabla_\nu \mu}{\mu^2} - \nabla^\nu \omega \nabla_\nu \omega\right) + \frac{3}{G} V_G(G) - V'_G(G) + G\left[\frac{V_\mu(\mu)}{\mu^2} + V_\omega(\omega)\right] + \frac{G}{16\pi}(R + 2\Lambda) = 0, \tag{12}$$

$$\left(\frac{2\nabla_\alpha G \nabla_\beta G}{G^2} - \frac{\nabla_\alpha \nabla_\beta G}{G}\right)(g^{\alpha\beta} g_{\mu\nu} - \delta^\alpha_\mu \delta^\beta_\nu) - 8\pi\left[\left(\frac{1}{4\pi} G \omega \mu^2 \phi_\alpha \phi_\beta - \frac{\partial_\alpha G \partial_\beta G}{G^2} - \frac{\partial_\alpha \mu \partial_\beta \mu}{\mu^2} + \partial_\alpha \omega \partial_\beta \omega\right) \times \left(\delta^\alpha_\mu \delta^\beta_\nu - \frac{1}{2} g^{\alpha\beta} g_{\mu\nu}\right) + \frac{1}{4\pi} G \omega \left(B^\alpha{}_\mu B_{\nu\alpha} + \frac{1}{4} g_{\mu\nu} B^{\alpha\beta} B_{\alpha\beta}\right) + g_{\mu\nu}\left(\frac{1}{4\pi} G V_\phi(\phi) + \frac{V_G(G)}{G^2} + \frac{V_\mu(\mu)}{\mu^2} + V_\omega(\omega)\right)\right] + R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + g_{\mu\nu} \Lambda = -8\pi G T_{\mu\nu}. \tag{13}$$

4. STATIC, SPHERICALLY SYMMETRIC VACUUM SOLUTION

In the static, spherically symmetric case with line element

$$ds^2 = Bdt^2 - Adr^2 - r^2 d\Omega^2, \tag{14}$$

and with $d\Omega^2 = d\theta^2 + \sin^2\theta\, d\phi^2$, the field equations are written as

$$\frac{1}{A}\mu^2\phi_r + \frac{\partial V_\phi}{\partial \phi_r} = \frac{4\pi}{A\omega}J_r, \tag{15}$$

$$\phi_t'' + \frac{2}{r}\phi_t' + \frac{\omega'}{\omega}\phi_t' + \frac{1}{2}\left(3\frac{A'}{A} - \frac{B'}{B}\right)\phi_t' - A\mu^2\phi_t + AB\frac{\partial V_\phi}{\partial \phi_t} = -\frac{4\pi A}{\omega}J_t, \tag{16}$$

$$G'' + \frac{2}{r}G' - \frac{3}{2}\frac{G'^2}{G} + \frac{1}{2}\left(\frac{\mu'^2}{\mu^2} - \omega'^2\right)G + \frac{1}{2}\left(\frac{B'}{B} - \frac{A'}{A}\right)G' + AV_G'(G) - 3A\frac{V_G(G)}{G} - AG\left[\frac{V_\mu(\mu)}{\mu^2} + V_\omega(\omega)\right] - \frac{AG(R+2\Lambda)}{16\pi} = 0, \tag{17}$$

$$\mu'' + \frac{2}{r}\mu' - \frac{\mu'^2}{\mu} - \frac{G'}{G}\mu' + \frac{1}{4\pi}G\omega\left(\phi_r^2 - \frac{A}{B}\phi_t^2\right)\mu^3 + \frac{1}{2}\left(\frac{B'}{B} - \frac{A'}{A}\right)\mu' - 2A\frac{V_\mu(\mu)}{\mu} + AV_\mu'(\mu) = 0, \tag{18}$$

$$\omega'' + \frac{2}{r}\omega' - \frac{G'}{G}\omega' + \frac{1}{8\pi}G\mu^2\left(\frac{A}{B}\phi_t^2 - \phi_r^2\right) + \frac{1}{2}\left(\frac{B'}{B} - \frac{A'}{A}\right)\omega' + \frac{1}{8\pi B}G\phi_t'^2 - \frac{1}{4\pi}AGV_\phi(\phi) - AV_\omega'(\omega) = 0, \tag{19}$$

$$8\pi G T_t^t = -\Lambda - V - \frac{1}{A}N + \frac{A'}{A^2 r} - \frac{1}{Ar^2} + \frac{1}{r^2} + \frac{G''}{AG} + \frac{2}{r}\frac{G'}{AG} - 2\frac{G'^2}{AG^2} - \frac{1}{2}\frac{A'G'}{A^2G} - \omega G\left(\frac{\phi_t'^2}{AB} + \frac{\mu^2\phi_t^2}{B} + \frac{\mu^2\phi_r^2}{A}\right), \tag{20}$$

$$8\pi G T_r^r = -\Lambda - V + \frac{1}{A}N - \frac{B'}{ABr} - \frac{1}{Ar^2} + \frac{1}{r^2} + \frac{1}{2}\frac{B'G'}{ABG} + \frac{2}{r}\frac{G'}{AG} - \omega G\left(\frac{\phi_t'^2}{AB} - \frac{\mu^2\phi_t^2}{B} - \frac{\mu^2\phi_r^2}{A}\right), \tag{21}$$

$$8\pi G T_r^t = -2\frac{G\omega\mu^2\phi_t\phi_r}{B}, \qquad 8\pi G T_t^r = 2\frac{G\omega\mu^2\phi_t\phi_r}{A}, \tag{22}$$

$$\begin{aligned} 8\pi G T_\theta^\theta = 8\pi G T_\phi^\phi = &-\Lambda - V - \frac{1}{A}N + \frac{1}{2}\frac{A'}{A^2 r} + \frac{1}{4}\frac{A'B'}{A^2 B} \\ &-\frac{1}{2}\frac{B'}{ABr} + \frac{1}{4}\frac{B'^2}{AB^2} - \frac{1}{2}\frac{B''}{AB} + \frac{G''}{AG} + \frac{1}{r}\frac{G'}{AG} - 2\frac{G'^2}{AG^2} - \frac{1}{2}\frac{A'G'}{A^2 G} \\ &+\frac{1}{2}\frac{B'G'}{ABG} - \omega G\left(-\frac{\phi_t'^2}{AB} - \frac{\mu^2\phi_t^2}{B} + \frac{\mu^2\phi_r^2}{A}\right), \end{aligned} \tag{23}$$

where

$$R = \frac{B''}{AB} - \frac{B'^2}{2AB^2} - \frac{A'B'}{2A^2 B} + \frac{2B'}{ABr} - \frac{2A'}{A^2 r} + \frac{2}{Ar^2} - \frac{2}{r^2}, \tag{24}$$

$$N = -4\pi\left(\frac{\mu'^2}{\mu^2} + \frac{G'^2}{G^2} - \omega'^2\right), \tag{25}$$

$$V = 2\omega G V_\phi(\phi) + 8\pi\left[\frac{V_G(G)}{G^2} + \frac{V_\mu(\mu)}{\mu^2} + V_\omega(\omega)\right]. \tag{26}$$

The prime denotes differentiation with respect to r, i.e., $y' = dy/dr$.

These equations can be substantially simplified in the matter vacuum case ($T_\nu^\mu = 0$), with no cosmological constant ($\Lambda = 0$), setting the potentials to zero ($V_\phi = V_G = V_\mu, V_\omega = 0$) and also setting $\phi_r = 0$. These choices leave us with six equations in the six unknown functions A, B, ϕ_t, G, μ, and ω:

$$\begin{aligned} &\frac{B'G'}{2ABG} - \frac{G'}{AGr} + 2\omega G\left(\frac{\phi_t'^2}{AB} + \frac{\mu^2\phi_t^2}{B}\right) - \frac{B''}{2AB} + \frac{B'^2}{4AB^2} \\ &\qquad + \frac{A'B'}{4A^2 B} - \frac{B'}{2ABr} - \frac{A'}{2A^2 r} + \frac{1}{Ar^2} - \frac{1}{r^2} = 0, \end{aligned} \tag{27}$$

$$\begin{aligned} &\frac{G''}{AG} - \frac{2G'^2}{AG^2} - \frac{B'G'}{2ABG} - \frac{A'G'}{2A^2 G} + \frac{B'}{ABr} + \frac{A'}{A^2 r} \\ &+8\pi\left(\frac{G'^2}{AG^2} + \frac{\mu'^2}{A\mu^2} - \frac{\omega'^2}{A} - \frac{\omega G\mu^2\phi_t^2}{4\pi B}\right) = 0, \end{aligned} \tag{28}$$

$$\begin{aligned} &\omega G\left(\frac{\phi_t'^2}{AB} - \frac{\mu^2\phi_t^2}{B}\right) + 4\pi\left(\frac{G'^2}{AG^2} + \frac{\mu'^2}{A\mu^2} - \frac{\omega'^2}{A}\right) \\ &\qquad + \frac{B'G'}{2ABG} + \frac{2G'}{AGr} - \frac{B'}{ABr} - \frac{1}{Ar^2} + \frac{1}{r^2} = 0, \end{aligned} \tag{29}$$

$$\mu'' + \frac{2}{r}\mu' - \frac{\mu'^2}{\mu} - \frac{G'}{G}\mu' + \frac{1}{2}\left(\frac{B'}{B} - \frac{A'}{A}\right)\mu' - \frac{A\omega G\phi_t^2}{4\pi B}\mu^3 = 0, \tag{30}$$

$$\omega'' + \frac{2}{r}\omega' - \frac{G'}{G}\omega' + \frac{1}{2}\left(\frac{B'}{B} - \frac{A'}{A}\right)\omega' + \frac{G}{2B}\phi_t'^2 + \frac{AG\mu^2\phi_t^2}{8\pi B} = 0, \tag{31}$$

$$G'' + \frac{2}{r}G' - \frac{3}{2}\frac{G'^2}{G} + \frac{1}{2}\left(\frac{B'}{B} - \frac{A'}{A}\right)G' + \frac{1}{2}\left(\frac{\mu'^2}{\mu^2} - \omega'^2\right)G - \frac{1}{16\pi}AGR = 0, \tag{32}$$

The values of A, B, and B' are fixed by the requirement that at large distance from a source, we must be able to mimic the Schwarzschild solution (albeit with a modified gravitational constant), and that at spatial infinity, the metric must be asymptotically Minkowskian. The vector field ϕ must also vanish at infinity, which provides another boundary condition. Next, we assume that the values of G, μ, and ω are dependent on the source mass only, i.e., $G' = \mu' = \omega' = 0$. We seek the remaining initial conditions in the form of the fifth force charge Q_5, and initial values of $G = G_0$, $\mu = \mu_0$, and $\omega = \omega_0$. We note that the basic properties of the numerical solution and the solution's stability are not affected by the values chosen for these parameters. However, their values must be chosen such that they correctly reflect specific physical situations. To determine these values, we now turn to the case of the point test particle.

5. TEST PARTICLE EQUATION OF MOTION

We begin by defining a test particle via its Lagrangian:

$$\mathscr{L}_{\rm TP} = -m + \alpha\omega q_5 \phi_\mu u^\mu, \tag{33}$$

where m is the test particle mass, α is a factor representing the nonlinearity of the theory (to be determined later), ω is present as it determines the interaction strength, q_5 is the test particle's fifth-force charge, and $u^\mu = dx^\mu/ds$ is its four-velocity.

We assume that the test particle charge is proportional to its mass:

$$q_5 = \kappa m, \tag{34}$$

with κ constant and independent of m. This assumption implies that the fifth force charge q_5 is not conserved, as mass is not conserved. This is the case in Maxwell-Proca theory, as $\nabla^\mu J_\mu \neq 0$. We also have that the fifth force source charge $Q_5 \propto M$.

From (33), the equation of motion is obtained

$$m\left(\frac{du^\mu}{ds} + \Gamma^\mu_{\alpha\beta}u^\alpha u^\beta\right) = -\alpha\kappa\omega m B^\mu{}_\nu u^\nu. \tag{35}$$

That m cancels out of this equation is nothing less than a manifestation of the equivalence principle.

Our acceleration law can be written as [5]:

$$\ddot{r} = -\frac{G_N M}{r^2}\left[1 + \alpha - \alpha(1+\mu r)e^{-\mu r}\right], \tag{36}$$

where G_N is Newton's gravitational constant and α is given by

$$\alpha = \frac{M}{(\sqrt{M}+E)^2}\left(\frac{G_\infty}{G_N}-1\right), \tag{37}$$

where E is a constant of integration.

The acceleration law (36) can also be recast in the commonly used Yukawa form:

$$\ddot{r} = -\frac{G_Y M}{r^2}\left[1+\alpha_Y\left(1+\frac{r}{\lambda}\right)e^{-r/\lambda}\right], \tag{38}$$

with the Yukawa parameters α_Y and λ given by

$$G_Y = \frac{G_N}{1+\alpha_Y}, \tag{39}$$

$$\alpha_Y = -\frac{(G_\infty - G_N)M}{(G_\infty - G_N)M + G_N(\sqrt{M}+E)^2}, \tag{40}$$

$$\lambda = 1/\mu = \frac{\sqrt{M}}{D}. \tag{41}$$

Here, E and D are two universal constants of integration which can be determined from fits to galaxy rotation curve data.

We can also express the acceleration law (36) as

$$\ddot{r} = -\frac{G_{\rm eff} M}{r^2}, \tag{42}$$

where the effective gravitational constant $G_{\rm eff}$ is defined as

$$G_{\rm eff} = G_N\left[1+\alpha-\alpha(1+\mu r)e^{-\mu r}\right]. \tag{43}$$

The metric parameter $B(r)$ is given by

$$B(r) = 1 - \frac{2G_N M}{r} + \frac{(1+\alpha)G_N^2 M^2}{r^2}. \tag{44}$$

The $B(r)$ and $A(r)$ solutions are shown in Figure 1.

6. COSMOLOGY

In the case of a homogeneous, isotropic cosmology, using the Friedmann-Lemaître-Robertson-Walker (FLRW) line element,

$$ds^2 = dt^2 - a^2(t)[(1-kr^2)^{-1}dr^2 + r^2 d\Omega^2], \tag{45}$$

the field equations assume the following form:

$$\ddot{\mu} + 3H\dot{\mu} - \frac{\dot{\mu}^2}{\mu} - \frac{\dot{G}}{G}\dot{\mu} + \frac{1}{4\pi}G\omega\mu^3\phi_0^2 + \frac{2}{\mu}V_\mu - V_\mu' = 0, \tag{46}$$

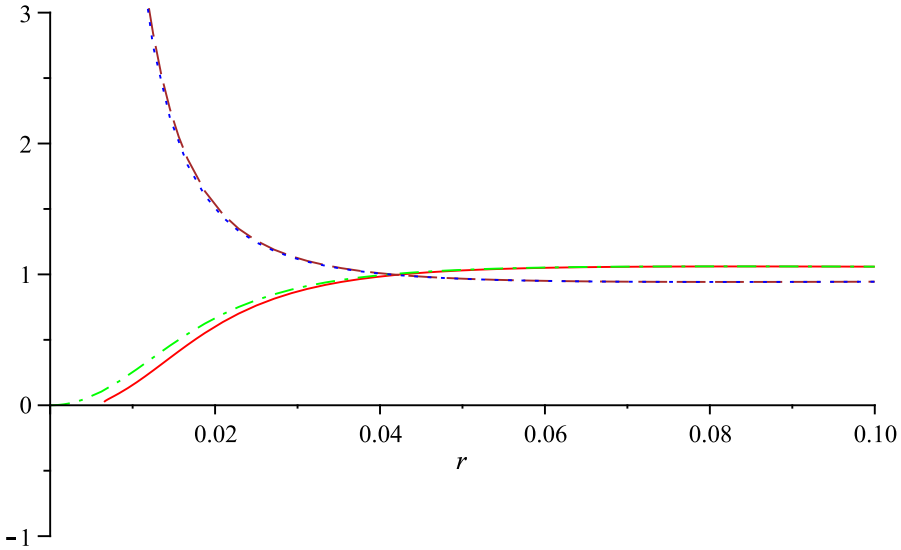


FIGURE 1. Comparing MOG numerical solutions to the Reissner-Nordström solution, for a $10^{11}\ M_\odot$ source mass. The MOG metric parameters A (solid red line) and B (dashed brown line) are plotted along with the Reissner-Nordström values of A (dash-dot green line) and B (dotted blue line). Horizontal axis is in pc. We observe that the A metric parameter reaches 0 at below the Schwarzschild radius of a $10^{11}\ M_\odot$ mass, which is ~ 0.01 pc.

$$\ddot{\omega} + 3H\dot{\omega} - \frac{\dot{G}}{G}\dot{\omega} - \frac{1}{8\pi}G\mu^2\phi_0^2 + \frac{1}{4\pi}GV_\phi + V'_\omega = 0, \tag{47}$$

$$\ddot{G} + 3H\dot{G} - \frac{3}{2}\frac{\dot{G}^2}{G} + \frac{G}{2}\left(\frac{\dot{\mu}^2}{\mu^2} - \dot{\omega}^2\right) + \frac{3}{G}V_G - V'_G + G\left[\frac{V_\mu}{\mu^2} + V_\omega\right] + \frac{G}{8\pi}\Lambda - \frac{3G}{8\pi}\left(\frac{\ddot{a}}{a} + H^2\right) = 0, \tag{48}$$

$$H^2 + \frac{k}{a^2} = \frac{8\pi G\rho}{3} - \frac{4\pi}{3}\left(\frac{\dot{G}^2}{G^2} + \frac{\dot{\mu}^2}{\mu^2} - \dot{\omega}^2 - \frac{1}{4\pi}G\omega\mu^2\phi_0^2\right) + \frac{2}{3}\omega GV_\phi + \frac{8\pi}{3}\left(\frac{V_G}{G^2} + \frac{V_\mu}{\mu^2} + V_\omega\right) + \frac{\Lambda}{3} + H\frac{\dot{G}}{G}, \tag{49}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{8\pi}{3}\left(\frac{\dot{G}^2}{G^2} + \frac{\dot{\mu}^2}{\mu^2} - \dot{\omega}^2 - \frac{1}{4\pi}G\omega\mu^2\phi_0^2\right) + \frac{2}{3}\omega GV_\phi + \frac{8\pi}{3}\left(\frac{V_G}{G^2} + \frac{V_\mu}{\mu^2} + V_\omega\right) + \frac{\Lambda}{3} + H\frac{\dot{G}}{2G} + \frac{\ddot{G}}{2G} - \frac{\dot{G}^2}{G^2}, \tag{50}$$

$$\omega\mu^2\phi_0 - \omega\frac{\partial V_\phi}{\partial\phi_0} = 4\pi J_0, \qquad J_i = 0, \tag{51}$$

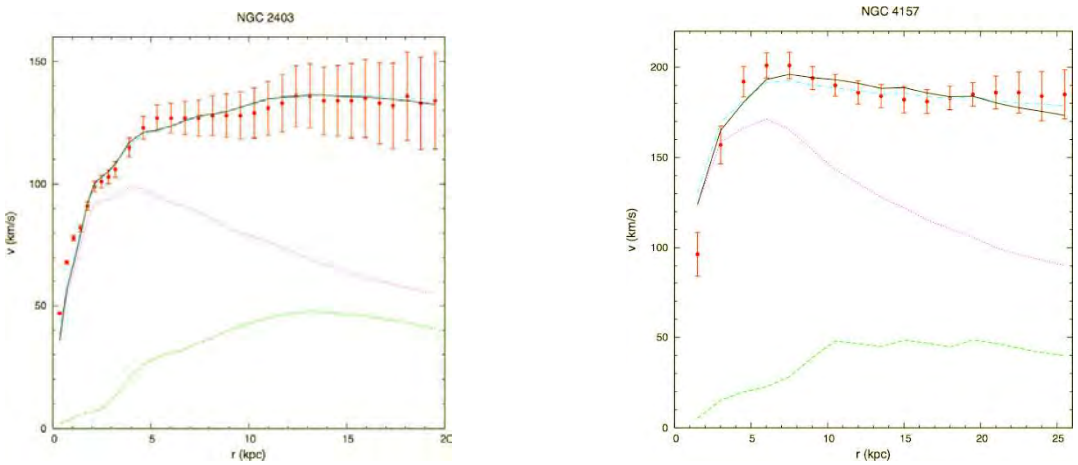


FIGURE 2. Photometric fits to galaxy rotation curves. There are 2 benchmark galaxies presented here. Each is a best fit via the single parameter (M/L) based on the photometric data of the gaseous (HI plus He) and luminous stellar disks. The radial coordinate (horizontal axis) is given in kpc and the rotational velocity (vertical axis) in km/s. The red points with error bars are the observations, the solid black line is the rotation curve determined from MOG, and the dash-dotted cyan line is the rotation curve determined from MOND [7]. The other curves are the Newtonian rotation curves of the various separate components: the long-dashed green line is the rotation curve of the gaseous disk (HI plus He) and the dotted magenta curve is that of the luminous stellar disk (from [8, 9].

where $H = \dot{a}/a$ is the Hubble expansion rate.

It is possible to obtain an exact numerical solution to this set of equations using numerical methods [5, 6]. To carry out the solution, we assume a pressureless matter equation of state $w = p/\rho = 0$. Detailed fits to cosmological data, including the CMB angular power spectrum, the matter power spectrum and the SN1a supernovae data have been obtained [6]. We find that the solutions can yield a "bouncing" cosmology. The bounce can be fine-tuned by choosing an appropriate value for V_G. This ensures that the universe reaches sufficient density in order to form a surface of last scattering. We emphasize that in our model only ordinary baryonic matter is present with a matter density of $\sim$5% of the critical density. Nevertheless, the cosmology is flat, due in part to the increased value of the gravitational constant G, and in part to the presence of the non-zero energy density associated with V_G.

7. FITTING GALAXY, CLUSTER DATA AND SOLAR SYSTEM DATA

Unless one assumes that a massive dark matter halo is present, a typical spiral galaxy is dominated in mass by the central bulge. The motion of stars in the outer reaches of a galaxy can, therefore, be well approximated by the equations of motion in a static, spherically symmetric vacuum field. Indeed, our experience shows that the flat rotation curves of galaxies provide a sensitive test to determine the values of the constants D and E. In particular, it is easy to see that our results so far are compatible with the Tully-Fisher law [10].

Kepler's laws of orbital motion yield a relationship between circular orbital velocity v_c at radius r from a mass M in the form

$$\frac{v_c^2}{r} = \frac{GM}{r^2}. \tag{52}$$

Tully and Fisher [10] have determined that for galaxies, assuming that the brightness of a galaxy and its mass are correlated, the flat part of the rotation curve obeys the empirical relationship:

$$v_c^n \propto M, \tag{53}$$

where $3 \lesssim n \lesssim 4$. In our case, we obtain

$$v_c^2 \propto \sqrt{M}, \tag{54}$$

corresponding to $n = 4$ in the Tully-Fisher relationship.

Taking the next step, we select a small sample of galaxies and obtain an approximate fit to these galaxies yielding the values

$$D \simeq 6250\, M_\odot^{1/2} \mathrm{kpc}^{-1}, \tag{55}$$

$$E \simeq 25000\, M_\odot^{1/2}. \tag{56}$$

The galaxy rotation curves we obtain for galaxies of varying mass are in good agreement with these values, treating D and E as universal constants without dark matter (Figure 2). The galaxy rotation curves were obtained modeling the galaxies as point masses, benefiting from photometric data, as in the more extensive fit to galaxy rotation velocities [8, 9, 11]. This exercise demonstrates that our established relationships between M, α, and μ not only satisfy the Tully-Fisher relationship, but also offer good agreement with actual observations. N-body simulations of galaxy rotation curve dynamics using MOG have been performed [12].

In [9, 13], the spherically symmetric, static vacuum solution was used successfully to model galaxy clusters. We are able to produce a comparable result, while keeping the parameters D and E constant, by introducing an additional assumption: that the values of the MOG parameters G_∞ and μ at some distance r from the center of a spherically symmetric distribution of matter are determined not only by the amount of matter contained within radius r, but by the amount of matter within radius r^*. Figure 3 shows the case of $r^* = 3r$.

We have also succeeded in fitting the bullet cluster data [15], using MOG without dark matter [16, 17].

We have applied MOG to predict dispersion velocity curves for globular clusters, and found that the predictions follow those of Newtonian gravity [18]. By using Sloan Digital Sky Survey (SDSS) data, we have investigated how modified gravity theories including MOND and MOG affect satellite galaxies with the result that the data cannot currently differentiate significantly between modified gravity theories and dark matter models [19]. The MOG prediction for lensing caused by intermediate galaxies and clusters of galaxies has been investigated [20].

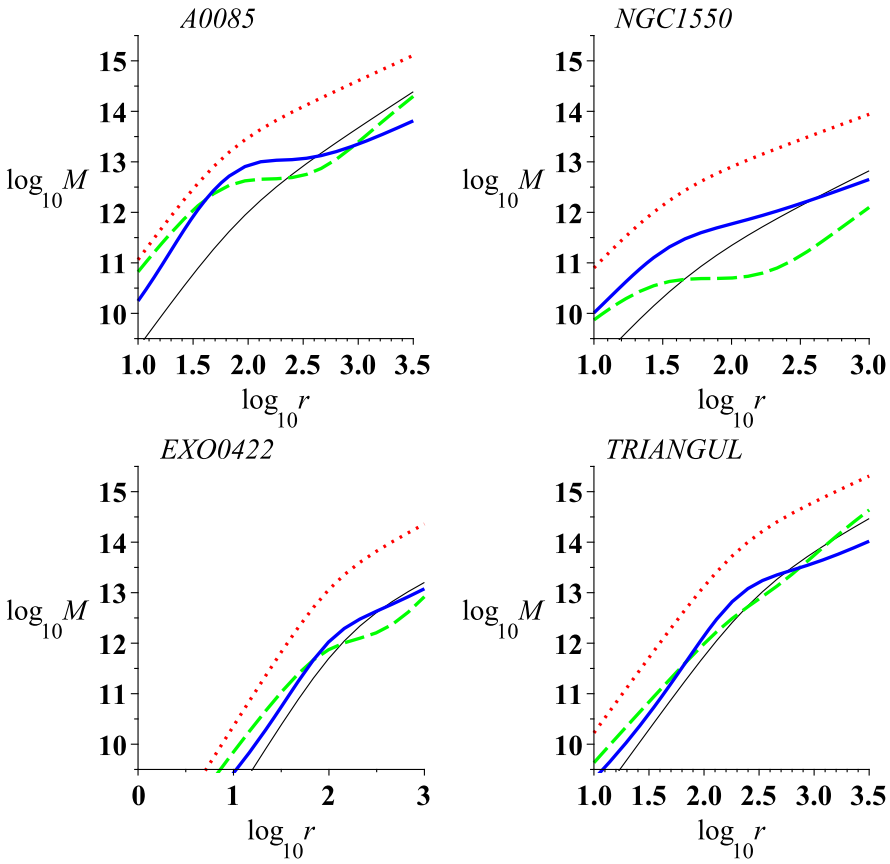


FIGURE 3. A small sample of galaxy clusters studied in [9, 13]. Thin (black) solid line is the mass profile estimate from [14]. Thick (blue) solid line is the mass profile estimated using our STVG results. Dashed (green) line is the result published in [13], while the dotted (red) line is the Newtonian mass profile estimate. Radial distances are measured in kpc, masses in $M_\odot$.

The theory must also be consistent with experiments performed within the solar system or in Earthbound laboratories. Several studies (see, e.g., [21]) have placed stringent limits on Yukawa-like modifications of gravity based on planetary observations, radar and laser ranging, and other gravity experiments. However, our prediction of the absolute value of the α_Y parameter is very small when λ_Y is small. The latter is estimated at $\lambda_Y \simeq 0.16$ pc ($\sim 5 \times 10^{15}$ m, or about 33,000 AU) for the Sun, and $\lambda_Y \simeq 2.8 \times 10^{-4}$ pc ($\sim 8.7 \times 10^{12}$ m, or ~ 58 AU) for the Earth. The corresponding values of $|\alpha_Y|$ are $|\alpha_Y| \simeq 3 \times 10^{-8}$ and $|\alpha_Y| \simeq 9 \times 10^{-14}$, respectively, clearly not in contradiction with even the most accurate experiments to date (Figure 4).

In the solar system the MOG field equations become essentially those of the Jordan-Brans-Dicke model [22, 23], for the influence of the vector field ϕ is reduced to very small values as shown in Figure 4. However, the standard JBD model coupling constant $\omega_{\rm JBD}$ has to be fine-tuned $\omega_{\rm JBD} > 40,000$ to fit the Cassini spacecraft measurement of the Eddington-Robertson, parameterized post-Newtonian parameter $\gamma - 1 = (2.1 \pm 2.3) \times 10^{-5}$; the other parameter β satisfies $\beta = 1$ in MOG. We have resolved this problem in MOG by coupling the scalar field G directly to matter by means of a scalar matter

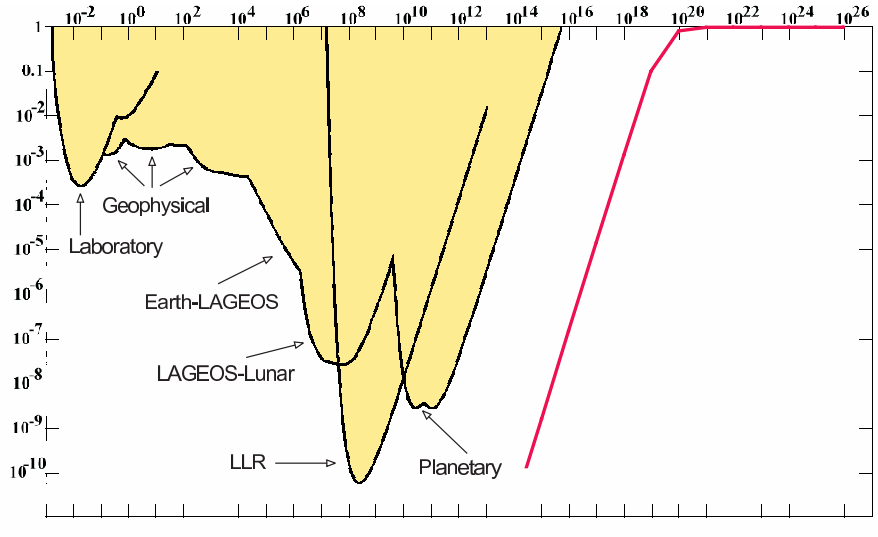


FIGURE 4. Predictions of the Yukawa-parameters from the MOG field equations are not in violation of solar system and laboratory constraints. Predicted values of λ (horizontal axis, in m) vs. $|\alpha_Y|$ are indicated by the solid red line. Plot adapted from [21].

current:

$$J = -\frac{1}{2}GT, \tag{57}$$

where $T = T^{\mu}{}_{\mu}$. This leads to obtaining an agreement with Earth based equivalence principle experiments and $\gamma = 1$ [24].

We have plotted M vs. $r_0 = \mu^{-1}$ in Figure 5. For the purposes of this plot, we used previously published results, while noting that our new calculations place dwarf galaxies, galaxies, and galaxy clusters by definition exactly on the line representing our prediction. This plot demonstrates the validity of MOG from the scales of star clusters to cosmological scales.

We have investigated the possibility that MOG can explain in a fundamental way the origin of inertial mass. The static, spherically symmetric solution does not satisfy Birkhoff's theorem as in the case of the Schwarzschild solution in GR. This leads to a Mach-type influence of distant matter that can determine the inertial mass of a body. A possible spacecraft experiment has been proposed to test this prediction [25].

On the scale of Earth-based laboratory and solar system experiments with ever greater precision, MOG may eventually be verified or falsified. Beyond the solar system, as larger galaxy samples become available, the presence or absence of baryonic oscillations in the matter power spectrum may unambiguously decide in favor of modified gravity theories or dark matter [5, 6]. Confirmed detection of dark matter particles in deep space or in the laboratory would also be a strong indication against modified gravity.

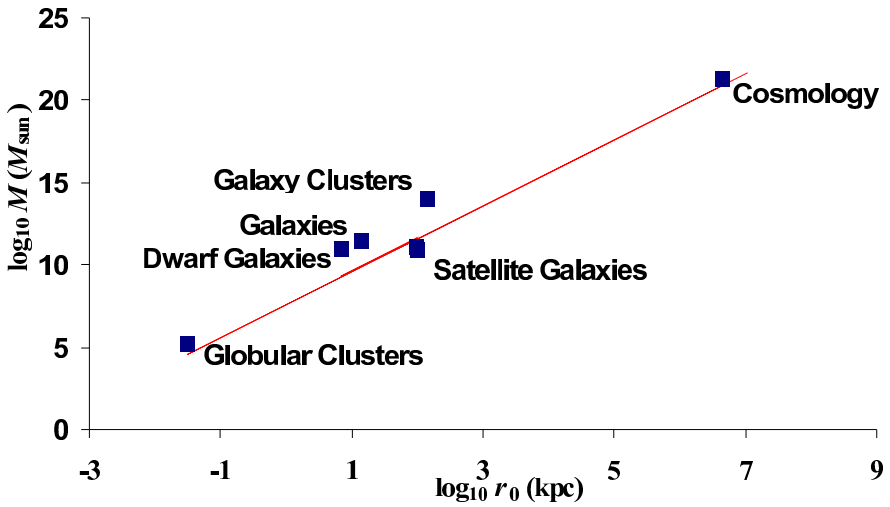


FIGURE 5. The relationship $\mu^2 M = \text{const.}$ between mass M and the Yukawa-parameter $r_0 = \mu^{-1}$ across many orders of magnitude remains valid. The solid red line represents our theoretical prediction.

8. CONCLUSIONS

In this paper, we have demonstrated how results of our Modified Gravity (MOG) theory can be derived directly from the action principle, without resorting to the use of fitted parameters. After we fix the values of some integration constants from observations, no free adjustable parameters remain, yet the theory remains consistent with observational data in the two cases that we examined: the vacuum solution of a static, spherically symmetric gravitational field, and a cosmological solution. These solutions were explored using numerical methods, avoiding the necessity to drop terms or make other simplifying assumptions in order to obtain an analytic solution. Further, the constraints used to compute the solutions are consistent to the extent that they overlap with one another. The fact that at the level of the calculations presented here, our theory is not obviously falsified is an indication that we should pursue MOG further, for instance by obtaining interior solutions to the MOG field equations, and using these solutions to develop tools to perform N-body simulations.

ACKNOWLEDGMENTS

The research was partially supported by National Research Council of Canada. Research at the Perimeter Institute for Theoretical Physics is supported by the Government of Canada through NSERC and by the Province of Ontario through the Ministry of Research and Innovation (MRI).

REFERENCES

1. E. Komatsu et al., arXiv:1001.4538 [astro-ph].
2. S. Perlmutter, Astroph. J., **517**, 565 (1999).
3. A. G. Riess et al., Astrophys. J., 116, 1009, 1998.
4. J. W. Moffat, JCAP(03):004, 2006.
5. J. W. Moffat and V. T. Toth, Class. Quant. Grav. **26** 085002, 2009.
6. V. T. Toth, arXiv:1011.5174 [gr-qc].
7. M. Milgrom, Astrophys. J. **270**, 365 (1983); Astrophys. J. **270**, 371 (1983).
8. J. R. Brownstein and J. W. Moffat, Astrophys. J., **636**, 721Ű741, 2006.
9. J. R. Brownstein, arXiv:0908.0040 [astro-ph].
10. R. B. Tully and J. R. Fisher, A&A, **54**, 661, 1977.
11. D. C. Rodrigues, P. S. Letelier, and I. L. Shapiro, JCAP0404020, 2010.
12. C. S. S. Brandao and J. C. N. de Araujo, Astrophs. J. **717**, 849 (2010), arXiv:1006.1000.
13. J. R. Brownstein and J. W. Moffat, Mon. Not. R. Astron. Soc., **367**, 527, 2006.
14. T. H. Reiprich and H. BÍohringer, Astrophys. J., **567**,716, 2002.
15. D. Clowe, S. W. Randall, and M. Markevitch, Nucl. Phys. Proc. Suppl. **173**, 28,2007.
16. J. R. Brownstein and J. W. Moffat, Mon. Not. R. Astron. Soc., **382**, 29, 2007.
17. J. W. Moffat and V. T. Toth, arXiv:1005.2685 [astro-ph].
18. J.W. Moffat and V. T. Toth, Astrophys. J., **680**, 1158, 2008.
19. J. W. Moffat and V. T. Toth, arXiv:0901.1927 [astro-ph].
20. J. W. Moffat and V. T. Toth, Mon. Not. Roy. Astron. Soc., **397**, 1885, 2009.
21. E. G. Adelberger, B. R. Heckel, and A. E. Nelson, Ann. Rev. of Nucl. and Part. Sc, **53**, 77, 2003.
22. P. Jordan. Schwerkraft und Weltall, Grundlagen der Theoretische Kosmologie. Vieweg und Sohn, Braunschweig, 1955.
23. C. Brans and R. H. Dicke, Phys. Rev., **124**, 925, 1962.
24. J. W. Moffat and V. T. Toth, arXiv:1001.1564 [gr-qc].
25. J. W. Moffat and V. T. Toth, Mon. Not. Roy. Astron. Soc., **395**, L25-L28, 2009.

Galaxy Rotation Curves from General Relativity with Infrared Renormalization Group Effects

Davi C. Rodrigues[*], Patricio S. Letelier[†] and Ilya L. Shapiro[**]

[*]*Departamento de Física, CCE, Universidade Federal do Espírito Santo, 29075-910, Vitória, ES, Brazil.*
[†]*Departamento de Matemática Aplicada, IMECC, Universidade Estadual de Campinas, 13083-859, Campinas, SP, Brazil.*
[**]*Departamento de Física, ICE, Universidade Federal de Juiz de Fora, 36036-330, Juiz de Fora, MG, Brazil.*

Abstract. We review our contribution to infrared Renormalization Group (RG) effects to General Relativity in the context of galaxies. Considering the effective action approach to Quantum Field Theory in curved background, we argued that the proper RG energy scale, in the weak field limit, should be related to the Newtonian potential. In the galaxy context, even without dark matter, this led to a remarkably small gravitational coupling G variation (about or less than 10^{-12} of its value per light-year), while also capable of generating galaxy rotation curves about as good as the best phenomenological dark matter profiles (considering both the rotation curve shape and the expected mass-to-light ratios). Here we also comment on related developments, open issues and perspectives.

INTRODUCTION

Currently there is a large body of data coming from cosmological and astrophysical observations that is mostly consistent with the existence of dark matter. Such observations also suggest that the hypothesized particles that constitute dark matter have very small cross section and travel at speeds much lower than light. These lead to the cold dark matter framework, which is one of the pillars of the current standard cosmological model ΛCDM.

It is not only tempting, but mandatory to check if such dark matter particles exist (by detecting them in laboratory based experiments, for instance) and also to check if the gravitational effects that lead to the dark matter hypothesis could follow from a more detailed and complete approach to gravity. The effects of pure classical General Relativity at galaxies have been studied for a long time and, considering galaxy rotation curves, the differences between General Relativity and Newtonian gravity are negligible, since in a galaxy matter moves at speeds much lower than that of light and is typically subject to weak gravitational fields ($\Phi \ll c^2$), which leads to the Newtonian limit of General Relativity [1].

[1] There are some proposals that consider General Relativity in the context of galaxies which do not lead to Newtonian gravity, see e.g. [1, 2, 3, 4, 5]. It is not impossible that a reasonable explanation for galaxy rotation curves may rely on similar approaches, nevertheless up to now none of such proposals have found

There is however a newer approach that may change considerably the role of dark matter, while following standard physical principles. Namely, the investigation of the running of the gravitational coupling parameter G on large scales as induced by the renormalization group framework.

The running of coupling constants is a well known phenomenon within Quantum Field Theory. It is well-known that the renormalization group method can be extended to quantum field theory on curved space time and to some models of quantum gravity (see, e.g., [6]), such that the beta functions can be interpreted in this framework. Concerning the high energy (UV) behavior, there is hope that the running of G in quantum gravity may converge to a non-Gaussian fixed point (asymptotic safety) [7, 8]. Our present concern is, however, not about the UV completeness, but with the behavior of G in the far infrared regime (IR). In the electromagnetic case the IR behavior corresponds to the Appelquist-Carazzone decoupling [9] (see e.g., [10] for a recent derivation of this theorem). In the case of gravity the same effect of decoupling has been obtained in [11, 12], but only for the higher derivative terms in the gravitational action. It remains unclear whether such decoupling takes place for the other terms. This possibility was studied on phenomenological grounds a number of times before, e.g. [13, 14].

In [15] we presented new results on the application of renormalization group (RG) corrections to General Relativity (GR) in the astrophysical domain. Previous attempts to apply this picture to galaxies have considered for simplicity point-like galaxies. We extended previous considerations by identifying the proper renormalization group energy scale μ and by evaluating the consequences considering the observational data of disk galaxies. Also we propose a natural choice for the identification of μ, linking it to the local value of the Newtonian gravitational potential. With this choice, the renormalization group-based approach is capable to mimic dark matter effects with great precision. This picture induces a very small variation on the gravitational coupling parameter G, namely a variation of about 10^{-7} of its value across 10^5 light-years. We call our model RGGR, in reference to renormalization group effects in General Relativity.

In order to evaluate the observational consequences of the RGGR model and to compare it to other proposals, recent high quality observational data [16, 17] from nine regular spiral galaxies were mass-modelled using the standard procedures for the baryonic part, and four different models for the "dark" component: i) the RGGR model; ii) one of the most phenomenological successful dark matter profiles, the Isothermal profile [18]; and two alternative models which were built to avoid the need for the dark matter: iii) the Modified Newtonian Dynamics (MOND) [19, 20] and iv) the Scalar-Tensor-Vector Gravity (STVG) [21]. The latter is a recent relativistic proposal that is capable of dealing with galaxy rotation curves and other phenomena usually attributed to dark matter. For galaxy rotation curves phenomena, STVG becomes equivalent to a similar proposal called MSTG [22, 23, 24]. The model parameters that we use to fit galaxies in the STVG framework can be found in ref. [23]. While for MOND we use the a_0 value as given in [25].

The quality of the rotation curve fits and total stellar mass as inferred from the RGGR model is perfectly satisfactory considering both the general behavior of the model and

---

a baryonic mass distribution that is in conformity with astrophysical expectations.

its results when applied to nine particular galaxies, as analyzed in [15]. It is about the same of the Isothermal profile quality, while it seems significantly better than the quality of MOND and STVG. In the case of MOND, we did numerical experiences with a_0 as a free parameter and found that, albeit the concordance with the shape of the rotation curve can considerably increase in this case, the concordance with the expected stellar mass-to-light ratios remains unsatisfactory (similar conclusions have also appeared in some recent papers, e.g. [17, 26, 27], and it seems that the concordance can only be improved by adjusting the MOND's $\mu(x)$ function in an *ad-hoc* way).

THE RUNNING OF G

The β-function for the gravitational coupling parameter G has been discussed in the framework of different approaches to Quantum Gravity and Quantum Field Theory in curved space-time. In [15] we followed the derivation used previously in [13]. If G does not behave as a constant in the far IR limit, it was argued in [13] (and recently in more details in [28]) that the logarithmic running of G is a direct consequence of covariance and must hold in all loop orders. As far as direct derivation of the physical running of G is not available, it is worthwhile to explore the possibility of a logarithmically running G at the phenomenological level.

Consider the following infrared β-function for General Relativity,

$$\beta_{G^{-1}} \equiv \mu \frac{dG^{-1}}{d\mu} = 2\nu \frac{M^2_{\text{Planck}}}{c\hbar} = 2\nu G_0^{-1}. \tag{1}$$

Equation (1) leads to the logarithmically varying $G(\mu)$ function,

$$G(\mu) = \frac{G_0}{1 + \nu \ln(\mu^2/\mu_0^2)}, \tag{2}$$

where μ_0 is a reference scale introduced such that $G(\mu_0) = G_0$. The constant G_0 is the gravitational constant as measured in the Solar System (actually, there is no need to be very precise on where G assumes the value of G_0, due to the smallness of the variation of G). The dimensionless constant ν is a phenomenological parameter which depends on the details of the quantum theory leading to eq. (2). Since we have no means to compute the latter from first principles, its value should be fixed from observations. It will be shown that even a very small ν can lead to observational consequences at galactic scales.

The action for this model is simply the Einstein-Hilbert one in which G appears inside the integral, namely,

$$S_{\text{RGGR}}[g] = \frac{c^3}{16\pi} \int \frac{R}{G} \sqrt{-g}\, d^4x. \tag{3}$$

In the above, G should be understood as an external scalar field that satisfies (2). Since for the problem of galaxy rotation curves the cosmological constant effects are negligible, we have not written the Λ term above. Nevertheless, for a complete cosmological picture, Λ is necessary and it also runs covariantly with the RG flow of G (see e.g.,[13]).

There is a simple procedure to map solutions from the Einstein equations with the gravitational constant G_0 into RGGR solutions. One need not to follow this route, one may find all the dynamics from the RGGR equations of motion, which can be found by a direct variation of the action (3) in respect to the metric, leading to equations of motion that have the same form of those of a scalar-tensor gravity[2]. In this review, we will proceed to find RGGR solutions via a conformal transformation of the Einstein-Hilbert action, and to this end first we write

$$G = G_0 + \delta G, \tag{4}$$

and we assume $\delta G/G_0 \ll 1$, which will be justified latter. Introducing the conformally related metric

$$\bar{g}_{\mu\nu} \equiv \frac{G_0}{G} g_{\mu\nu}, \tag{5}$$

the RGGR action can be written as

$$S_{\text{RGGR}}[g] = S_{\text{EH}}[\bar{g}] + O(\delta G^2), \tag{6}$$

where S_{EH} is the Einstein-Hilbert action with G_0 as the gravitational constant. The above suggest that the RGGR solutions can be generated from the Einstein equations solutions via the conformal transformation (5). Indeed, within a good approximation, one can check that this relation persists when comparing the RGGR equations of motion to the Einstein equations even in the presence of matter [15].

In the context of rotation curves of galaxies, standard General Relativity gives essentially the same predictions of Newtonian gravity. The Newtonian potential Φ_{Newt} is related to the metric by

$$\bar{g}_{00} = -\left(1 + \frac{2\Phi_{\text{Newt}}}{c^2}\right). \tag{7}$$

Hence, using eq. (5), the effective RGGR potential Φ in the non-relativistic limit is given by

$$\Phi = \Phi_{\text{Newt}} + \frac{c^2}{2}\frac{\delta G}{G_0}. \tag{8}$$

An equivalent result can also be found from the evaluation of a test particle geodesics [15]. In the context of weak gravitational fields $\Phi_{\text{Newt}}/c^2 \ll 1$ (with $\Phi_{\text{Newt}} = 0$ at spatial infinity) holds, and hence the term $\delta G/G_0$ should not be neglected.

In order to derive a test particle acceleration, we have to specify the proper energy scale μ for the problem setting in question, which is a time-independent gravitational phenomena in the weak field limit. This is a recent area of exploration of the renormalization group application, where the usual procedures for high energy scattering of particles cannot be applied straightforwardly. Previously to [15] the selection of $\mu \propto 1/r$, where r is the distance from a massive point, was repeatedly used, e.g. [29, 30, 31, 32, 13]. This identification adds a constant velocity proportional to ν to

[2] We stress that it is only the from since RGGR is not a type of scalar-tensor gravity, and G is not a fundamental field of the model.

any rotation curve. Although it was pointed as an advantage due to the generation of "flat rotation curves" for galaxies, it introduced difficulties with the Tully-Fisher law, the Newtonian limit, and the behavior of the galaxy rotation curve close to the galactic center, since there the behavior is closer to the expected one without dark matter. In [15] we introduced a μ identification that seems better justified both from the theoretical and observational points of view. The characteristic weak-field gravitational energy does not comes from the geometric scaling $1/r$, but from the Newtonian potential Φ_{Newt}. However, the straight relation $\mu \propto \Phi_{\text{Newt}}$ leads to $\mu \propto 1/r$ in the large r limit; which is unsatisfactory on observational grounds (bad Newtonian limit and correspondence to the Tully-Fisher law). One way to recover the Newtonian limit is to impose a suitable cut-off, but this does not solves the Tully-Fisher issues [13]. Another one is to use [15]

$$\frac{\mu}{\mu_0} = \left(\frac{\Phi_{\text{Newt}}}{\Phi_0}\right)^{\alpha}, \tag{9}$$

where Φ_0 and α are constants. Apart from the condition $0 < \Phi_0 < c^2$ (i.e., essentially Φ_0 is a reference Newtonian potential), the precise value of Φ_0 is largely irrelevant for the problem of rotation curves. The relevant parameter is α. It is a phenomenological parameter that depends on the mass of the system, and it must go to zero when the mass of the system goes to zero. This is necessary to have a good Newtonian limit. From the Tully-Fisher law, it is expected to increase monotonically with the increase of the mass. Such behavior is indeed found from the galaxy fits done in [15]. In a recent paper, an upper bound on $\nu\alpha$ in the Solar System was derived [28]. In galaxy systems, $\nu\alpha|_{\text{Galaxy}} \sim 10^{-7}$, while for the Solar System, whose mass is about 10^{-10} of that of a galaxy, $\nu\alpha|_{\text{Solar System}} \lesssim 10^{-17}$. It shows that a linear decrease on α with the mass is sufficient to satisfy both the current upper bound from the Solar System and the results from galaxies.

We also point that the above energy scale setting (9) was recently re-obtained from a more theoretical perspective [33].

Once the μ identification is set, it is straightforward to find the rotation velocity for a static gravitational system sustained by its centripetal acceleration,

$$V^2_{\text{RGGR}} \approx V^2_{\text{Newt}} \left(1 - \frac{\nu\alpha c^2}{\Phi_{\text{Newt}}}\right). \tag{10}$$

Contrary to Newtonian gravity, the value of the Newtonian potential at a given point does play a significant role in this approach. This sounds odd from the perspective of Newtonian gravity, but this is not so from the General Relativity viewpoint, since the latter has no free zero point of energy. In particular, the Schwarzschild solution is not invariant under a constant shift of the potential.

In the following, we will comment on the effect of the relation (10) to galaxy rotation curves. First from a more general perspective, and then to the modeling of individual galaxies.

GALAXY ROTATION CURVES

Before proceeding to specific galaxy rotation curves modeling, it is more instructive to analyze general features of the relation (10), and to compare it to the standard approach. In [15] we analyze some general aspects and scaling laws of the RGGR model, with no dark matter, in comparison to the isothermal profile; both of them, at this step, without gas and with an exponential stellar disk. In particular, it was pointed that the RGGR rotation curves have a reasonable shape to fit galaxies (i.e., no clear problems like increasing or decreasing too fast, oscillations...), and that they effectively behave similarly to cored dark matter profiles at inner radii, whose effective core radius scales with the galaxy disk scale length. Further details in our paper.

We have also extended the previous analysis by adding a gas-like contribution (a re-scaled version of the NGC 3198 gaseous part). In particular, this numerically evaluates how the model behaves on the presence of density perturbations at large radii. In the first plot of fig. (1) it is displayed the result for RGGR, which is remarkably good (a similar plot can be found in our original paper), while in the others plots in fig. (1) (presented at the Conference, but not in [15]) one sees the results for the same mass distribution but different choices for the energy scale[3] μ.

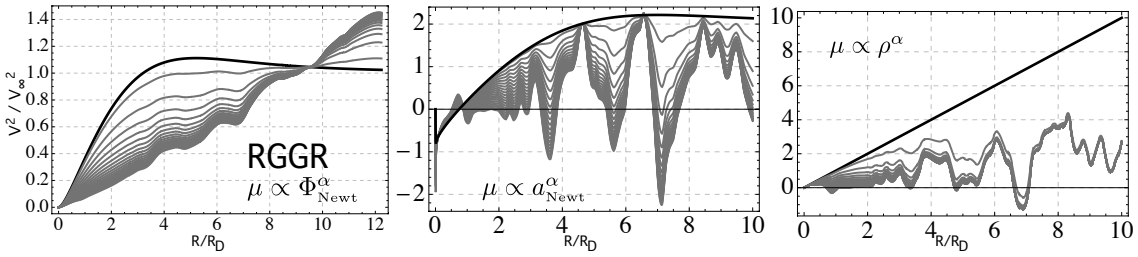


FIGURE 1. The additional circular velocity squared induced by different choices of μ. The first plot refers to RGGR, the two others to different identifications of μ: one depends on the Newtonian acceleration (a variation inspired on MOND) and the other on the (baryonic) matter density. All the plots above display the additional squared velocity of each model divided by $V_\infty^2 \equiv \nu\alpha c^2$ and as a function of R/R_D, where R is the the radial cylindrical coordinate and R_D is the stellar disk scale length. Black lines depict the additional velocity due to a pure exponential stellar disk, while the gray solid lines take into account the gas mass $M_{\rm gas}$ for different values of $f \equiv M_{\rm gas}/M_{\rm stars}$, with $f = 0.2, 0.7, 1.2, ..., 9.7$ (i.e., the black lines stand for $f = 0$). See [15] for further details.

From fig. (1), the two other proposals different from RGGR are seen to be unsuited as replacements for dark matter. In particular, both are too sensitive to the gas presence, and both eventually add negative contributions to the total circular velocity at large radii.

In [15] we used a sample of nine high quality and regular rotation curves of disk galaxies from [16, 17]. In figs. (2, 3) we show one of ours results (see [15] for the complete set and further details) in comparison to the results of three other models: a cored dark matter profile (Isothermal profile), the Modified Newtonian Dynamics (MOND) and the recently proposed Scalar-Tensor-Vector Gravity (STVG).

---

[3] These other choices are also unsatisfactory from the theoretical perspective, since they have no direct relation to the local energy of the gravitational field in the weak field regime ($\Phi_{\rm Newt} \ll c^2$).

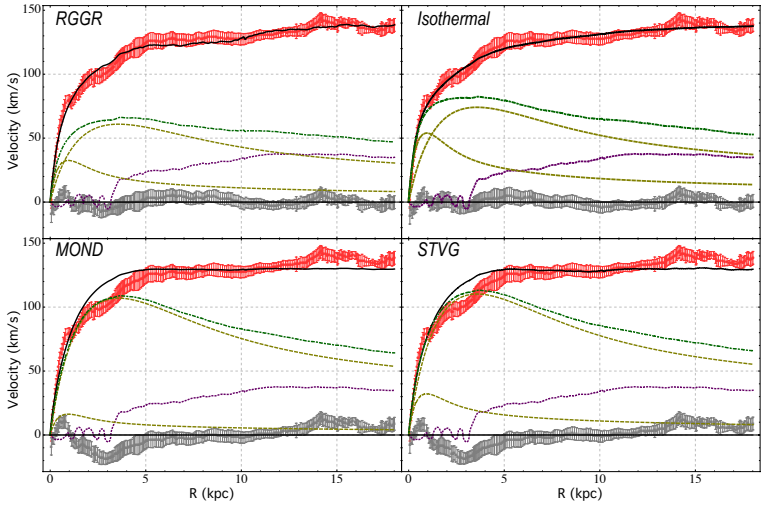


FIGURE 2. NGC 2403 rotation curve fits. The red dots and its error bars are the rotation curve observational data, the gray ones close to the abscissa are the residues of the fit. The solid black line for each model is its best fit rotation curve, the dashed yellow curves are the stellar rotation curves from the bulge and disk components, the dotted purple curve is the gas rotation curve, and the dot-dashed green curve is the resulting Newtonian, with no dark matter, rotation curve.

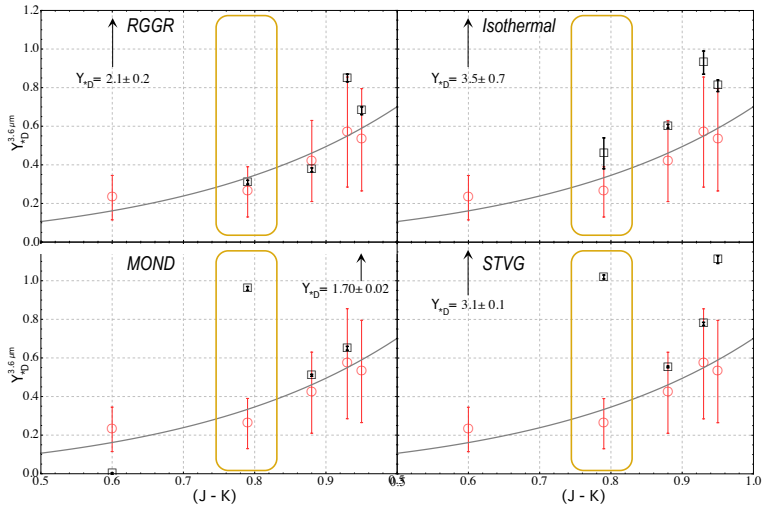


FIGURE 3. Stellar disk mass-to-light ratio (Y_{*D}) in the 3.6μm band as a function of the color $J-K$. Each galactic disk is represented above by an open circle, with a *reference* error bar of 50% of the Y_{*D} value. The black open squares display the Y_{*D} values and their associated 1σ errors for each galaxy as inferred from the rotation curve fits for each model. The highlighted square and circle correspond to the NGC 2403 galactic disk mass-to-light ratios. See [15] for further details.

Due to the considerably large uncertainty in the total stellar mass of each stellar component (disk and bulge), we first use the total stellar mass as a free parameter for the fittings (achieved from a χ^2 minimization considering the errors). At a second stage, we compare the resulting value with stellar population expectations, following the standard approach.

On the free parameters of each model, we remark that besides the total stellar mass, the Isothermal profile has two additional free parameters, the RGGR model has a single free parameter (α) while MOND and STVG have no free parameters that can vary from galaxy to galaxy. On the other hand, both of the latter depend on constants whose values are calibrated considering its best fit in a large sample of galaxies. We remark that the ν parameter in RGGR cannot vary from galaxy to galaxy, but α can, and galaxy rotation curves are sensible to the combination $\nu\alpha$, whose value is about the order of 10^{-7}. The best fit for NGC 2403 yields $\nu\alpha = (1.66 \pm 0.01) \times 10^{-7}$.

CONCLUSIONS

We presented a model, motivated by renormalization group corrections to the Einstein-Hilbert action, that introduces small inhomogeneities in the gravitational coupling across a galaxy (of about 1 part in 10^7) and can generate galaxy rotation curves in agreement with the observational data, without the introduction of dark matter as a new kind of matter. Both High and Low Surface Brightness galaxies were tested [15] . Considering the samples of galaxies evaluated in [15], the quality of the RGGR rotation curves, together with the corresponding mass-to-light ratios, is about the same than the Isothermal profile quality, but with one less free parameter. We expect that similar results would hold in regard to other cored dark matter profiles, while our results seem better than those achieved by the NFW profile [15]. We also compared the results of our model with MOND and STVG, and at face value our model yielded clearly better results.

Our results can be seen as a next step compared to the previous models motivated by renormalization group effects in gravity, e.g. [13, 34]. Their original analyses could only yield a rough estimate on the galaxy rotation curves, since they were restricted to modeling a galaxy as a single point. Trying to extend this approach to real galaxies, we have shown that the proper scale for the renormalization group phenomenology is not of a geometric type, like the inverse of the distance, but is related to the Newtonian potential with null boundary condition at infinity.

The essential feature for the RGGR rotation curves fits is the formula (10), which is by itself a simple formula that provides a very efficient description of galaxy rotation curves.

There are several tests and implications of this model yet to be evaluated. In particular we are working on applying the RGGR framework to a larger sample of galaxies (including elliptical galaxies) [35] and galaxy-galaxy strong lensing [36]. Related work on CMB, BAO and LSS in search for a new cosmological concordance model is also a work in progress [37].

Acknowledgements
DCR thanks FAPESP and PRPPG-UFES for partial financial support. PSL thanks CNPq and FAPESP for partial financial support. The work of I.Sh. was partially supported by CNPq, FAPEMIG, FAPES and ICTP.

REFERENCES

1. F. I. Cooperstock, and S. Tieu, *Int. J. Mod. Phys.*, **A22**, 2293–2325 (2007), astro-ph/0610370.
2. D. Vogt, and P. S. Letelier, *Mon. Not. Roy. Astron. Soc.*, **363**, 268–284 (2005), astro-ph/0507406.
3. F. I. Cooperstock, and S. Tieu, *Mod. Phys. Lett.*, **A23**, 1745–1755 (2008), 0712.0019.
4. D. Vogt, and P. S. Letelier (2005), astro-ph/0510750.
5. D. Vogt, and P. S. Letelier (2005), astro-ph/0512553.
6. I. L. Buchbinder, S. D. Odintsov, and I. L. Shapiro, *Effective action in quantum gravity*, Bristol, UK: IOP 413 p, 1992.
7. M. Niedermaier, and M. Reuter, *Living Rev. Rel.*, **9**, 5 (2006).
8. S. Weinberg, *PoS*, **CD09**, 001 (2009), 0908.1964.
9. T. Appelquist, and J. Carazzone, *Phys. Rev.*, **D11**, 2856 (1975).
10. B. Goncalves, G. de Berredo-Peixoto, and I. L. Shapiro, *Phys. Rev.*, **D80**, 104013 (2009), 0906.3837.
11. E. V. Gorbar, and I. L. Shapiro, *JHEP*, **02**, 021 (2003), hep-ph/0210388.
12. E. V. Gorbar, and I. L. Shapiro, *JHEP*, **06**, 004 (2003), hep-ph/0303124.
13. I. L. Shapiro, J. Sola, and H. Stefancic, *JCAP*, **0501**, 012 (2005), hep-ph/0410095.
14. M. Reuter, and H. Weyer, *JCAP*, **0412**, 001 (2004), hep-th/0410119.
15. D. C. Rodrigues, P. S. Letelier, and I. L. Shapiro, *JCAP*, **1004**, 020 (2010), 0911.4967.
16. W. J. G. de Blok, F. Walter, E. Brinks, C. Trachternach, S. Oh, and R. C. Kennicutt, *Astron. J.*, **136**, 2648–2719 (2008), 0810.2100.
17. G. Gentile, P. Salucci, U. Klein, D. Vergani, and P. Kalberla, *Mon. Not. Roy. Astron. Soc.*, **351**, 903 (2004), astro-ph/0403154.
18. K. G. Begeman, A. H. Broeils, and R. H. Sanders, *Mon. Not. Roy. Astron. Soc.*, **249**, 523–537 (1991).
19. M. Milgrom, *Astrophys. J.*, **270**, 365–370 (1983).
20. M. Milgrom, *Astrophys. J.*, **270**, 371–389 (1983).
21. J. W. Moffat, *JCAP*, **0603**, 004 (2006), gr-qc/0506021.
22. J. W. Moffat, *JCAP*, **0505**, 003 (2005), astro-ph/0412195.
23. J. R. Brownstein, and J. W. Moffat, *Astrophys. J.*, **636**, 721–741 (2006), astro-ph/0506370.
24. J. R. Brownstein (2009), 0908.0040.
25. R. H. Sanders, and S. S. McGaugh, *Ann. Rev. Astron. Astrophys.*, **40**, 263–317 (2002), astro-ph/0204521.
26. G. Gentile, B. Famaey, and W. J. G. de Blok (2010), 1011.4148.
27. B. Famaey, and J. Binney, *Mon. Not. Roy. Astron. Soc.*, **363**, 603–608 (2005), astro-ph/0506723.
28. C. Farina, W. J. M. Kort-Kamp, S. M. Filho, and I. L. Shapiro (2011), 1101.5611.
29. M. Reuter, and H. Weyer, *Phys. Rev.*, **D70**, 124028 (2004), hep-th/0410117.
30. D. A. R. Dalvit, and F. D. Mazzitelli, *Phys. Rev.*, **D50**, 1001–1009 (1994), gr-qc/9402003.
31. O. Bertolami, J. M. Mourao, and J. Perez-Mercader, *Phys. Lett.*, **B311**, 27–33 (1993).
32. J. T. Goldman, J. Perez-Mercader, F. Cooper, and M. M. Nieto, *Phys. Lett.*, **B281**, 219–224 (1992).
33. S. Domazet, and H. Stefancic (2010), 1010.3585.
34. M. Reuter, and H. Weyer, *Int. J. Mod. Phys.*, **D15**, 2011–2028 (2006), hep-th/0702051.
35. J. C. Fabris, P. L. C. Oliveira, D. C. Rodrigues, and I. L. Shapiro, work in progress.
36. D. C. Rodrigues, and P. S. Letelier, work in progress.
37. J. C. Fabris, I. L. Shapiro, and A. Toribio, work in progress.

PART II

GRAVITATIONAL DYNAMICS

OF GALAXY CLUSTERS

The Concordant Cosmological Model and the Dark Matter Hypothesis in Explaining Galaxy Cluster Dynamics

Hans Böhringer

1 Introduction

This talk will be quite observational and I'll give you first of all a description of what galaxy clusters are and how well we understand them. Then I will go on to talk about how we can test cosmological models in several ways. We can test cosmological models finding out whether their structure and the composition of the galaxy cluster is what we expect from theory. We can find out whether the number count of galaxy clusters comes out right, whether the cosmological models predict the right number of galaxy clusters for different parts of the redshift and then finally we can use galaxy clusters to trace their distribution on very large scales and look if that makes sense with the cosmological model.

2 Galaxy clusters

Galaxy clusters are actually an integral part of the large scale structure of the Universe and they form from density, positive density fluctuations with a large enough amplitude and a large enough mass that is collected in the galaxy cluster. Density fluctuations that are larger than the mean density of the Universe grow in the course of time. If they reach a threshold where the density is roughly about twice the critical density, this object, this mass agglomeration stops expanding with the Universe, collapses and forms objects. If you are interested in different masses of objects that have collapsed and have form, you can mathematically filter the fluctuation field. It's a filter that just contains the mass. The more you filter the field, the smoother it gets and the longer you'll have to wait until these filtered peaks reach a certain amplitude. If you look at the critical amplitude for the formation of objects today, you will find that the objects with a larger filter mass, that still reach this critical threshold to form objects, have masses of typically 1014 and 1015 solar masses and these are clusters of galaxies.

3 The Coma Cluster

Here we have a composite of two images of galaxy clusters. We have the optical image from the All Sky Survey and we see a large collection of galaxies (which is what gives the galaxy clusters their name); but they are much more than just a collection of galaxies. In X- ray images, for example of the All Sky Survey, performed by the Rossat satellite, you'll see a glowing of X-rays that extends over the whole cluster. This is a picture of the Coma cluster of galaxies [see image

below] at a distance of about 110 Mpcs from us and you see a hot gas that is shining in X-rays outlining the whole cluster. This already gives you a feeling. First of all, we see galaxy clusters as connected entities, as well defined objects. The big galaxies are typically in the cluster, but we have stars in the foreground and we have lots of galaxies. This already gives you a feeling that X-rays are probably a very nice tool to learn something about the structure and composition of galaxy clusters. Astronomers use a lot of X-ray observations to do that.

4 Hierarchy of structure of the Universe as defined by gravitational potential

To illustrate the significance of galaxy clusters in the hierarchy of the structure of the Universe, you can think of them as large mass associations that create deep gravitational potentials [see image below]. If you take a cross section through the potential of the Universe in this sketchy way, as shown over here, we will find galaxies which account for a gravitational potential. The depth of the potential is characterized by test particles that fill up this potential. The depth of the potential can be calculated or estimated from the velocity dispersion of the stuff that is in there or the gas that fills up the galaxy typically at high temperatures, for example, for our galaxy up to a million degrees. If you go to groups, they have a deeper potential. You have velocity dispersions up to 500 km per second for small galaxies running around the group and you have an X-ray temperature of the gas that is about a kilo-electron volt and then we go to really large galaxy clusters up to the most massive ones, which have temperatures in excess of 10 kilovolts and velocity dispersions of 1500 km/second. This really shows what well-defined objects galaxy clusters are and that we go into the realm of large structures in the Universe that are still determined by initial conditions. If we go to a super cluster, we don't have a deep gravitational potential, but lots of gravitational potentials of clusters and they have their signature at the same sort of velocity dispersion or temperature as a single galaxy cluster in our neighborhood.

5 Dark matter halos and galaxy clusters

One of the interesting things is the difference between characterizing this gravitational potential and characterizing the association of matter that some theoretical astrophysicists would call a dark matter halo [see image below]. If I stay in the conservative picture of Cold Dark matter (CDM), then a dark matter halo can directly be identified within a galaxy cluster. Looking at the galaxy cluster in X-ray, we see the whole dark matter halo and can characterize it. If we look at that dark matter halo associated to galaxies, it is very difficult to make the connection between the galaxy and the dark matter halo because the galaxy forms after gas was cooling and forming stars somewhere in the center of the halo. It doesn't tell you how much this dark matter halo extends and how much mass is contained, unless you do statistics like gravitational lensing, or look at satellite motions. So we have a very direct connection between dark matter halos and galaxy clusters. The only problem we have is that galaxy clusters are dynamically young and their dynamic time scale is of a few gigayears, not much less than Hubble time and we have constant growth of this dark matter halo by mergers, which disturbs the equilibrium structure of these objects. This is our largest problem to characterize the galaxy clusters: measure their mass when they are not in perfect equilibrium. Still we come from this theoretical side, what is the expectation if you're working with a CDM model and do numerical simulations? What do you expect for the structure of this dark matter halo coming out of these purely gravitational simulations? Here you have a nice picture [see image below] of how galaxy clusters form from

the assembly of smaller dark matter halos to form a big dark matter halo and you end up with something which at least in theory is described as a nice self-similar family of dark matter halos.

6 Distance and density scales

If you have an idea of how to scale different sizes of objects in relation to the same object, you can get them to line up. We scale them by looking at the different density distributions of dark matter halos, the small ones and the big ones. The amplitude, if they have just collapsed now, should be the same, but they have a larger size. You have the fiducial radius to scale their size. One way to do it is to find out whether the mean density of the cluster is just a certain multiple of the critical density of the Universe or the mean density of the Universe and you take this density radius as a scale radius for the cluster. If you then go back and scale these density profiles of different clusters that are forming in the simulations, you divide them by the same fiducial radius. These curves are falling nearly on top of each other, so we obtain a self-similar family of objects where the self similarity parameters are mass, formation time, and concentration parameter. We can scale the distance of different objects in the Universe as a function of redshift [see image below]. This is model-dependent. The whole picture changes even if you go from Einstein's model to the concordance cosmology model. You have to do everything correctly in each model and then test its consistency within the model. This was just a theoretical prelude to give you an idea of what to expect. We can measure mass profiles using the hydrostatic equation and at the end of this section I will come to how precisely we can actually apply that, but to some extent at least if we have relaxed clusters and we can assume that the gas is in hydrostatic equilibrium approximately in the galaxy clusters we form in this hydrostatic equation ... We get an equation that gives us the mass profile depending on several observables. The observables are the temperature profile in the cluster and the density profile in the cluster. The temperature profile absolutely, the density profile in a relative way. And apart from the assumption that the cluster is in hydrostatic equilibrium, in this way we also assume that the cluster is roughly symmetric. I will show this later.

7 X-ray observations

X-ray observations give us a possibility to deduce both the density and the temperature of the gas. We get the density, if you for example measure the X-ray luminosity in a certain bend, well a typical bend of an X-ray telescope... Luminosity is proportional to the emission that is measured, so the luminosity that we see is an integral of the line of sight density square of the gas roughly with a small temperature dependence that can be corrected. If you look at the spectrum, the spectrum is mostly formed by radiations with lines of some of the ions that have still kept some of the electrons ... It's a spectrum that comes from optical plasma with very simple atomic physics involved in it. If you use nuclear physics, which is relatively well understood, we can calculate what the spectrum is for some gas at the temperature of a few times 10 million Kelvin and we can calculate the continuum mostly from free radiations, which is how we can obtain the abundances of the elements as well as the temperature from the spectrum. One of the complications is that if you have several temperatures on the line sight, it's not always easy to unravel them. But let's assume that we get that done and then we obtain from the surface brightness profile of the galaxy cluster to the density profile. You measure the temperature from the spectra of several concentric rings of the X-ray emission in the cluster and we obtain this temperature profile. You have to de-project them and then get this mass profile of the cluster. These are clusters that have been selected because they look very regular.They are supposedly

quite relaxed and not recently disturbed from a major merger. So we obtain a mass scale, and the mass profile if you scale by the scaling rate, the fiducial radius, and if you scale by the total mass, we obtain the mass profile that has a shape which very well matches the Navarro, Frank, and White profile for these clusters. A similar result has been obtained by others from data that come from the Hassle peak X-ray observatory. So in this way we recover the density profile that can be predicted by numerical simulations. It's very difficult to distinguish between a Navarro, Frank and White and a Moore profile. In this way you can see all this relatively well, because it's difficult to get very good optical precision in the very center and at the outskirts.

8 Differences between the central parts of different galaxy clusters

In the central part of a regular cluster we have a central dominant galaxy. There is a lot of difficult physics going on in the center and it's very difficult to separate that from the details of the modeling. However, we have a very good consistency between model predictions and what we actually observe. We can extend that to clusters which are not regular. We have done a large survey on galaxy clusters that we selected only by their X-ray luminosity and redshift ,in a certain way, to optimally look at them with the XMM telescope. They go from small mass to large mass and are characterized by temperatures between 2 to 10 kilo-electrovolts for the galaxy clusters. We look at how much they can be explained as a self-similar family of objects. Again if we look at the density distribution of the gas determined from the surface brightness profile and you look at this scale radius and the scale density parameter over here, we see that we can get them to look very self-similar for a large part of the radius. This is a logarithmic scale for a large part of the radial range. But we have a problem in the inner part. If we look at the temperature profiles, we can see something similar. We see decreasing temperature profiles over most of the volume of the clusters and then we have this diversity in the inner part. This diversity comes from what was formerly called a cooling flow problem and now it is called a cool core problem. We have galaxy clusters with denser gas in the center and then usually they have a CD galaxy and a big black hole that fights back and then it shows feedback. We have galaxy clusters with less density and higher temperatures in the center which often have been recently disturbed by merges and they have less cool cores. So we have a high diversity here in the hydrodynamics, as far as what happens in the center of the clusters is concerned; but if we cut out this inner part, we have a very good feeling of how to characterize the statistics of this dark matter halo and to get a feeling of how well can determine mass, what scatter we have with different parameters.

9 Pressure profiles

One of the interesting parameters of course is the pressure because the derivative of the pressure is where we get the mass from and the pressure shows regular profiles if we scale it in the right way. And you can compare that also to model predictions of different types. Now, if we are going to hydrodynamics, this is much more complicated and we are not at the point where simulations know exactly what physical recipes they should put in because there are differences between different models. But you can see that there are three different types of simulations shown here. They are predicting these pressure profiles in these dark matter halos with differences of up to 20%. If you compare it to the observed pressure profiles, there is a relatively good equilibrium between the two so we are recovering the properties of galaxy clusters and get a good feeling for the statistics. They are a relatively nice self-similar family of course with certain variation that we try to characterize as statistical scatter in the relation.

10 Galaxy clusters and the mass of the dark matter halo

If you want to know how to use these galaxy clusters that we see for example in large surveys, that's one of the next steps I will be doing. We want to have a simple observable and then try to make a prediction for their mass or the mass of the dark matter halo. This is what theoreticians want to have to fit observations into their models. What we want to fit into the theoretical models is the mass of a dark matter halo within a certain fiducial radius and then we have a simple parameter that we can easily measure for a large number of clusters. For example, for X-ray luminosity, we can take these samples where we have deep X-ray observations and have measured the properties very well and,we can see that because of the self-similarity we get a very good correlation between the X-ray luminosity and the mass of the galaxy clusters. But we still have a scatter of more than 40% in predicting the mass, if you just take the X-ray luminosity. It gets better if you excise the center, where we have this diversity, and we can collapse the scatter of the relation by a good factor of two and get a much better prediction. X-ray measured gas mass is an even better predictor because it has a lower scatter, as low as sometimes 10% in predicting the mass.

11 Is it correct to apply hydrodynamic equilibrium to measure mass?

So we have now a very good understanding of the variety of galaxy cluster shapes and we have a good idea of how to predict their masses. The next thing of course you can now ask is: "Is it correct to apply hydrodynamic equilibrium to measure mass?" And you can also question if the spherical symmetry of the galaxy clusters is the right way to go. People have done hydrodynamic simulations to test how well these observational recipes work in determining the mass on theoretical clusters. I have to say that I still don't trust the hydrodynamics of simulations well enough and we can show in certain cases that there are differences between what we observe and what is modeled. But things come closer and closer together.

12 Gravitational lensing as another ways to measure mass

We also want to do it by observations. One of the ways, but still this is done with low statistics, is that we try to compare masses that are determined from gravitational lensing for galaxy clusters with the ones we get from X-rays, for example a sample of clusters of the Locus project. The X-ray analysis was done by Zhang. This is another set of observations from the Canadian team. Others too have published comparisons of the X-ray and lensing masses, for example Alexander Vikhlinin. The lensing measurements are made by the group of Hoekstra.

13 Measuring errors and bias

A problem here is that lensing masses have a large scatter, that deviates from the true probable scatter mass. One knows this from simulations, because you have a foreground structure and a background structure that is added to your lensing signal you will not, even in the best cases have a better estimate than 30% in measuring the mass. But for large samples it is believed that the lensing masses are unbiased while the X-ray masses could be biased in a certain way. So we need a lot of data to beat the scatter in this relation and to actually measure the bias between the two measurements. I think at the moment, taking about all the measurements we know, the best measurements, we see that probably the X-ray mass is too low an estimate, with a bias

of about 10% to 12% low. One of the reasons why the X-ray mass estimate is low is probably because what is not taken into account in the analysis that I have been showing is the motion of the gas and the gas has turbulent motion and slotting around. Hydrodynamic simulations show that this can be of the order of 10% to 20% of extra pressure that has not been taken into account in the mass estimate. So this is where we stand. I would say we have an idea of the galaxy cluster masses that we can determine from temperatures or luminosities in the best case with an accuracy of about10%. This is good enough to make nice measurements, to test cosmological models and to make a statement of Omega Matter, Omega Lambda, Sigma 8.

14 Components of galaxy cluster mass

Another interesting thing is that you can now look at how the galaxy clusters are, what are the components that make the mass of a galaxy cluster. We have several components. One is stellar mass in the galaxies, the next one is gas mass and there appears to be a large amount of missing mass –whatever it is, some say dark matter—, if we stay within the conventional concordance models using Newtonian gravity. One of the things we find is that the stellar mass fraction, the percentage of mass that is in the stars and in the galaxies is actually a decreasing function of the total mass of the system, but if you look at the gas mass, it is actually an increasing function of the system. The larger the cluster, the better it contains all the baryons inside and the gas against any feedback mechanism that is happening in the inside. The interesting thing is that if you add the two up and you compare the sum with the total mass value, in the conventional concordance model of baryonic mass, you need the dark matter mass. The distribution is [see image below]:

This means that stars and gas account for about 16% of the total galaxy cluster mass. This figure would approach 15%, when we take into account a little bit of loss of baryons for the most massive systems, but we still have this larg deficit for the less massive systems. We are not talking about dark energy, we're talking about matter, dark matter and baryons. There's a lot of interesting things happening here: a lot of hydrodynamics and very interesting astrophysics lots of baryons over here and in the stellar mass fractions, most of it still not perfectly understood. But I acknowledge the fact that for massive clusters, when one recovers the baryon fraction, one can even account for a little bit of the loss of baryons with the help of simulations.

15 Omega Baryonic and Dark Matter estimates

In a galaxy cluster we get 85% of dark matter and we get 15 percent of baryonic matter. If you want to see the agreement with the baryonic matter density obtained from Big Bang nucelo-synthesis, which is about 4 percent of the total density of the Universe (this is what we have for a Hubble constant of 70), you can come up with Omega Matter, which is less than the critical density of the Universe. That is the result that is about 15 years old, but could be derived from galaxy clusters. We see about 10 percent for the error bars. So, if we take the galaxy cluster composition as roughly representative for the Universe, we get a total matter density of something like 30 percent and a baryon density of about 4 percent comparing the cluster composition with nucleo-synthesis. Now you can do the next step and I'm a little bit skeptical on the application of that and the people who do it. I think the accuracy with which this is done is OK. They say, if we take the baryon fraction in clusters as an invariant, the measurement of this baryon fraction, the number that we obtain, depends on your cosmology and depends on the diameter distance. So doing the exercise and calculating baryon fractions for galaxy clusters at various redshifts within the frame of different cosmological models, you can get an almost invariant gas density.

This is the Einstein model, so you can use it to test your cosmology models. I used Steve Allen's results and tried to calculate the diameter distance from the X-gas. What they haven't done is that you can actually put these results in terms of a Hubble diagram and then you can see the diameter distance as a function of redshift [see image below] and you can see in a conventional way, as is the case with supernovae, which cosmological model best fits the Hubble diagram.. From a consistency check with different cosmological models they get an Omega Lambda within a certain range, and Omega Matter within a certain range. We are now getting the first few clusters at higher redshift. This is where reliable data are available. The scatter in the data and the statistics of the data just give you very interesting results on the Hubble constant, close to the value expected, but you would need more data and go to higher redshift here to distinguish between different cosmological models. So the preliminary result of this chapter is that we do find a good consistency in the shape of galaxy clusters, that we see roughly with what is predicted from cold dark matter models and the shape of dark matter halos. Using the galaxy clusters as cosmological yardsticks, assuming that the baryon fraction is an invariant, one gets the first set of cosmological tests.

16 Cluster mass functions and cosmological models

One of the questions in astrophysics is how far we can push for more precision because it's not easy to prove that the gas mass fraction is an invariant and it does rely a lot on hydrodynamics. But we can use galaxy cluster populations and not just galaxy clusters as cosmological probes. We can put these data into cosmological models. If one knows the underlying mass density distribution in the Universe, one can just predict which peaks in the mass distribution form clusters and from that one can predict cluster mass functions. This mass function has been tested with very large N-body simulations and has been modified. Such a function of mass starts with a mass function that has fewer objects with a high mass and then with time more and more massive objects are formed. If you do galaxy cluster counts as a function of redshift and as a function of mass you can actually show how the mass is growing with time and you can test cosmological models this way, because, how this mass function grows with time is very dependent on the cosmological model we use. We have used the Rossat Aall Sky Survey very intensively and I'm still working on that, still after 18 years. We have tried to identify as many galaxy clusters down to a certain down flux limit trying to obtain completeness, which is well over about 90% and then we obtain the most luminous clusters at different redshift shelves. At high redshifts, we only get the very luminous ones. Again the closer we come, the more luminous objects we have; but we have always had a complete sample above a limiting luminosity. Let me show you a lot of results which are part of our own Reflex survey with a lot of observations about and just about doubling the number for reflex 2. I'll show you the first results, but we also pushed the northern survey. So soon, in the next one or two years we'll increase the statistics of what I'm showing by a factor of 4. This shows a redshift distribution of the 400 galaxy clusters and if you have a good statistical eye you may see that this is not a homogeneous distribution, but that this distribution is clumped and we'll come back to that. If we have a cosmological model as indicated, we can predict how many clusters we see, but we can also predict how they are distributed in space. We have to choose the right cosmology. We have to make an assumption on the seed fluctuations, so we use the standard assumption of Harrison and double spectrum inflation, and we modify the power spectrum in the transfer function according to the type of matter you have and you put in baryonic acoustic oscillations, but they are so tiny for our purposes that they don't change the solutions very much. And then you can make these predictions and we have done that. We have used our galaxy clusters and determined statistically the X-ray luminosity function from

the known volume in which we count the clusters. The number count of clusters per unit volume as a function of X-ray luminosity and then we try to make predictions. We get the structure, we get the mass function of the galaxy cluster and then we use an empirical mass luminosity relation. You can get from the X-ray luminosity to the mass or from the predicted mass to the X-ray luminosity. Then, we plot on top of it the predicted X-ray luminosity function and the concordance cosmological model fits very nicely here. It fails for the groups, but in the group regime we don't have a good calibration for the mass and there is some cosmic variance problem in the sky which adds to that. This can probably be better done in one year from now. You see how little we know here and it is really difficult to find very good group samples to establish the mass calibration in this regime; but through trial and error we obtain two important cosmological parameters, the constraints for Omega Matter and for Sigma A. The sigma A is the amplitude scaling of the power spectrum. If we take the whole cluster survey we get something like Omega Matter density close to about 30% and the sigma A is about0.8%. While people were working on galaxy cluster cosmology, we defended the low sigma A for a long time when a lot of other people, gravitational lensing people and the first WMAP people wanted a higher sigma A. But we knew very well that we couldn't change sigma A that easily. Going from here to here you increase the number of galaxy clusters by more than a factor of two. We couldn't do that. We knew that and we were defending it until we were proven to be right. So we were quite proud with the results we had at the time.

17 The work of Vikhlinin on Dark Energy

Taking this result just for the present time, we cannot say much about dark energy or the cosmological constant. We just get a good constraint on Omega Matter, and you can combine it with supernova tests and WMAP and you can see the cross in the same region. Again this slide is several years old. Now you can go one step further and you can say I want to measure the mass function for the proxy mass function not at one epoch, but at different epochs and one of the nicest works that is available is the one of Alexei Vikhlinin of nearby galaxy clusters and their mass function. He has derived the mass function by using either the temperature or the gas mass, the total gas mass, as a proxy and from the calibration on what the mass should be he calculated the mass function and then he has a sample of galaxy clusters identified in Rossat pointed observations and at higher redshift and you can actually see how the mass function is changing. Going to a wrong cosmology, you don't reproduce these things but the right cosmology reproduces it. These two slides are shown by Alexei Vikhlinin to show exactly what you are always pointing out. You can only do a consistency test. You have to do all your calculations from the beginning for the right cosmology and then ask yourself whether it fits or not? You cannot start half way and then fit your results.

18 Error margins

Using this equation of a state parameter for dark energy which is somewhere with a large error bar of at least 10% or more, we obtain an Omega Lambda which is between 60% and 80%. So again, we have large error margins, but it fits very well with other cosmological constraints and some similar work has been done by the group of Steve Allen with Mantz. They have found similar constraints with large error margins scattering around Omega Matter between 0.1 and 0.4 and this is what you can get for small samples of galaxy clusters. If you want to do the next step and I mean... This is what everybody is aiming for. If you want to see how the cluster number density depends on the equation of the state parameter, even a variable stable equation of state

parameter, things become very tough. For example, if we just look at how many galaxy clusters we find as a function of redshift for a mass of about 1014 solar masses or 1013 solar masses you can see that changing the W parameter by about 30% gives you more than 30% difference in the number count. So it would be nice if we would have perfect masses for the galaxy clusters. It would not be too difficult to devise such a survey. But if one looks in more detail we can see that for example the number density changes with the mass calibration.... I mean a change of one percent in the mass scale gives you a 3Most of the populations are galaxy groups and they are much more difficult to handle than galaxy clusters. They are much more difficult to observe. And then it gets even tougher, so hopefully we can follow galaxy clusters with this future instrument from about a redshift of 2 to the present redshift, but that is one of the most interesting epochs if we talk about dark energy. Because increasing the redshift leverage is so important, we are doing a very elaborate program at the moment to try to find galaxy clusters with the X-ray observations that have been done by XMM and now we have about 30 clusters with nice X-ray emissions. They are very massive objects. It is the largest sample of distant galaxy clusters with high redshift and it is led in my group by René Fastbender, who has been mostly doing the work.

19 Far away massive galaxy clusters

Now one of the surprising things is that we started this in 2005 and we pretty much worked with photometry and spectroscopy so it was hard work, for more than 5 years, to find these 30 clusters with a redshift of about 0.8 and 17 clusters with a redshift of one. But we were rewarded very early to find a very unusual cluster at a redshift of 1.4. It turned out that now we have ages, the images that even the supernova people have followed up in several epochs. We have state age images from which you can do lensing and we have done a lensing analysis with the XMM and we have gathered data from which we have determined mass and it turns out to be a very massive object. If we then calculate what is the chance to find this mass, we find that it is small. Then again if you calculate how much volume you need to find these clusters at a redshift of 0.4, it would fit in the conventional concordance cosmology model. There are papers appearing in literature where people use modified gravity models and other things to explain these clusters, but we decide to be conservative and wait for more evidence. As a matter of fact, it is just one object, not enough to draw definite conclusions. We had a lot of internal discussion and we said we didn't want to draw conclusions from one object. We should continue and try to find more evidence, more should be found either by us or by other people. There are lots of opportunities to find them also in infrared surveys. So we have to wait, but I just wanted to say that there can always be surprises .....

20 Shücker's work

You can also do large scale structure work and I will be very quick with that. For the reflex survey we also measured the power spectrum. It was a paper done by Shuecker who unfortunately died in 2006. He was really one of our brains in this cosmological work that we did. He determined the power spectrum. He got three different boxes out of our survey and determined the conventional power spectrum. So there are three dependent data sets plotted on top of each other, but you can already see that this power spectrum doesn't dive down early enough to be consistent with the low omega model. So we have a power spectrum with more points in the low density realm compared to this one. These are the constraints that we obtain.

21 Conclusions

Galaxy clusters provide several independent and complementary means to measure or detect cosmological parameters. We can use the structures, number counts, and the large scale structure distribution and as far as our error margins are concerned, which are still substantial, we show that things are not perfect, but good enough to be compatible with the concordance cosmology model: So we don't really have any surprises yet. Maybe that one galaxy cluster we just mentioned, but we don't put that too high. I think a new epoch has come and I want to show you just one slide.

22 future research

And that is what we want to do. At our institute we're building a new instrument which is called e-ROSITA, an X-ray telescope that should be flying in a Russian satellite that is to be launched officially in 2012. Hopefully in 2013, it will do a four year sky survey. The main goal is to do cosmology with 50 thousand to 100 thousand clusters and increase the leverage we have now by a factor of 30. If we also do a lot of work to accompany it, we will also have perfect optical and gravitational lensing data and if we calibrate everything well, we will have some data to make a very good step in dark energy cosmology.

23 The composition of galaxy clusters

•78–87%= Dark Matter
• 11–14% = hot gas
• $2 - 6\%$ = galaxies (in total)

for$H_0 = 70$
H. Böhringer, Galaxien und Galaxienhaufen im Universum (§ 1) SS 2009

24 Galaxy Clusters Defined as Gravitational Potentials

X-ray emission originates from$20 - 100$ Mill. K plasma
$LX = 1043 - 31045erg/skT = 2 - 10keV$
ne $10 - 4 - 10 - 1cm - 3$

Figure 1: The Coma Galaxy Cluste (H. Böhringer, Galaxien und Galaxienhaufen im Universum (§ 1) SS 2009r)

Figure 2: Graph.-The Network of Large-Scale Structure is Formed by Many Zeldovich Pancakes The Coma cluster within the Great Wall, seen as a “finger of good” due to redshift space distortion

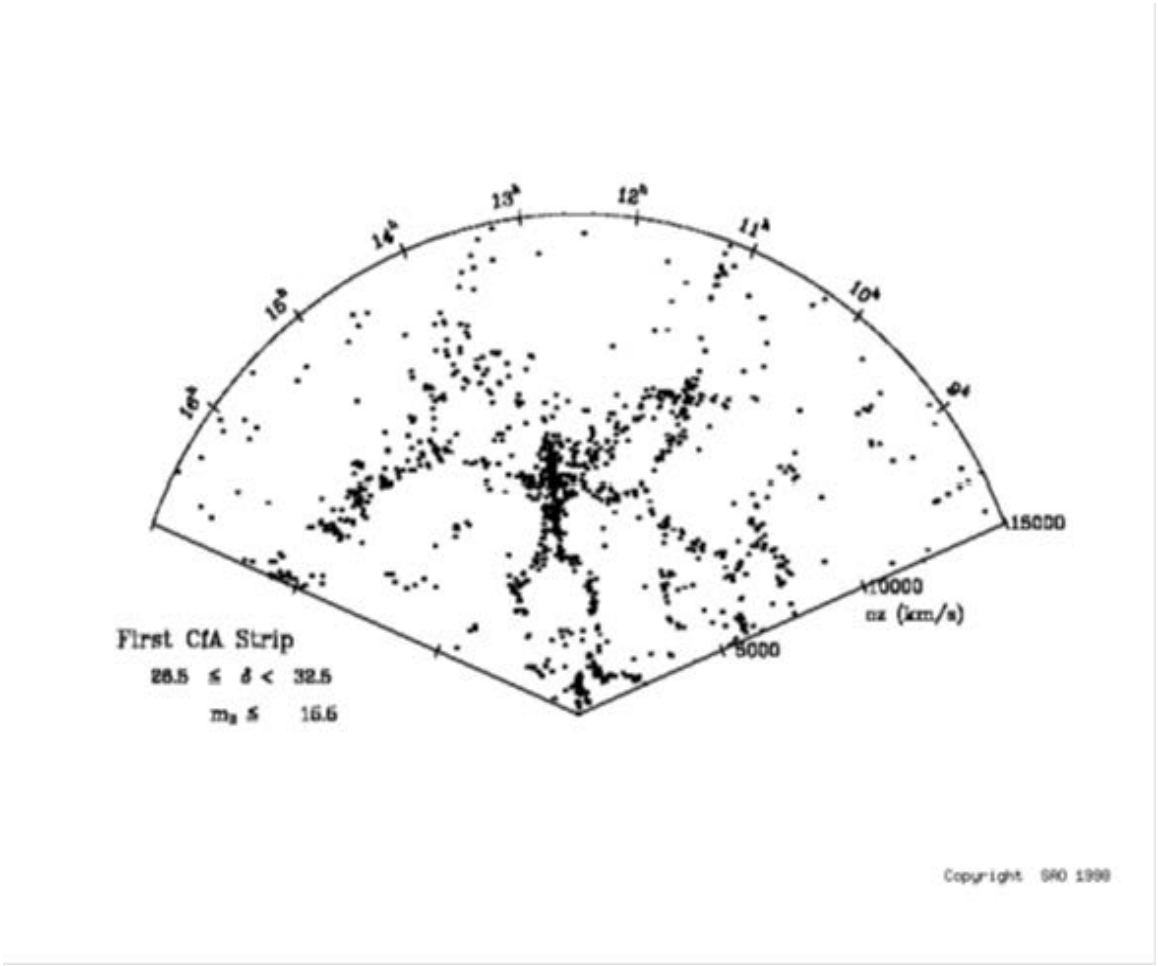


Figure 3: Sketch of the cosmic potential

Figure 4: Optical and X-ray appearance of the Coma cluster of galaxies (From POSS and ROSAT-All Sky Survey)

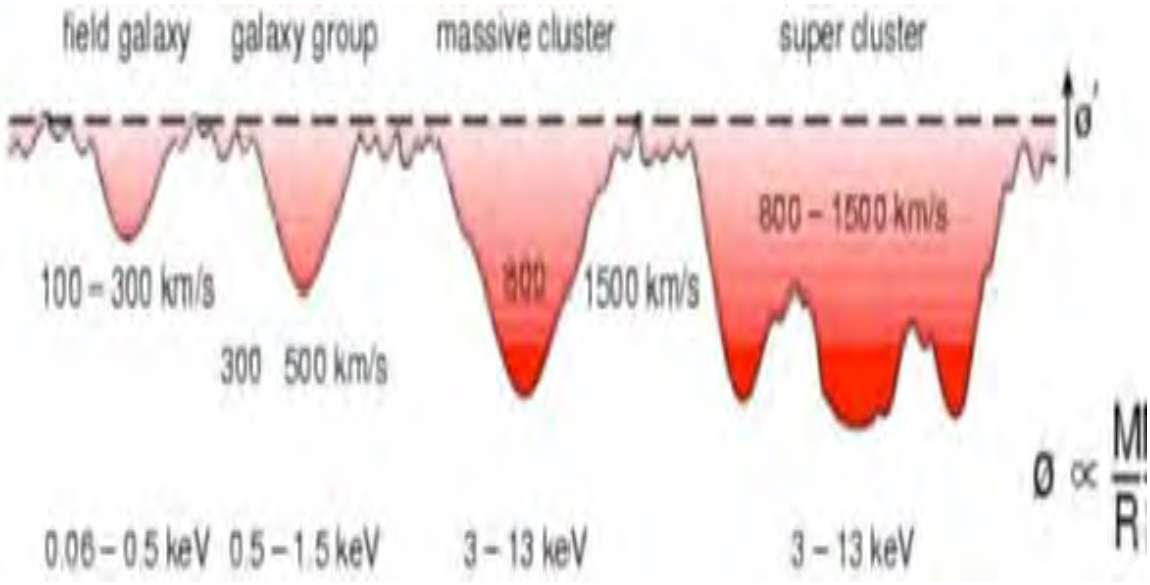


Figure 5: Comparison of galaxies and clusters as Dark Matter Halos
The intracluster gas is heated when the cluster forms and does not cool – it still reflects the potential depth (H. Böhringer, Galaxien und Galaxienhaufen im Universum (§ 1) SS 2009)

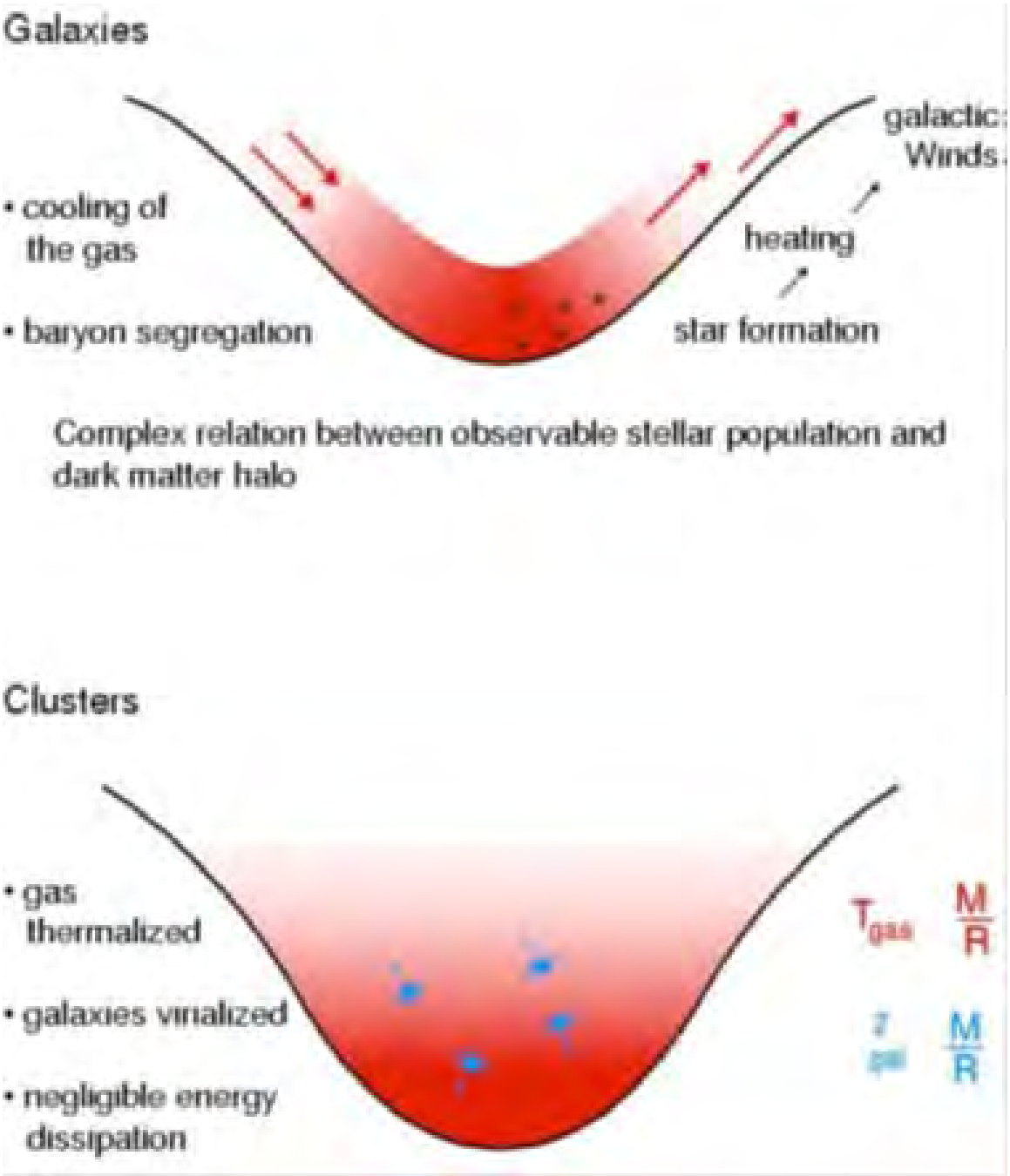


Figure 6: Merging Galaxy Clusters, illustrating how clusters of galaxies grow
The merging of two galaxy clusters in the system Abell 3528 (H. Böhringer, Galaxien und Galaxienhaufen im Universum (§ 1) SS 2009)

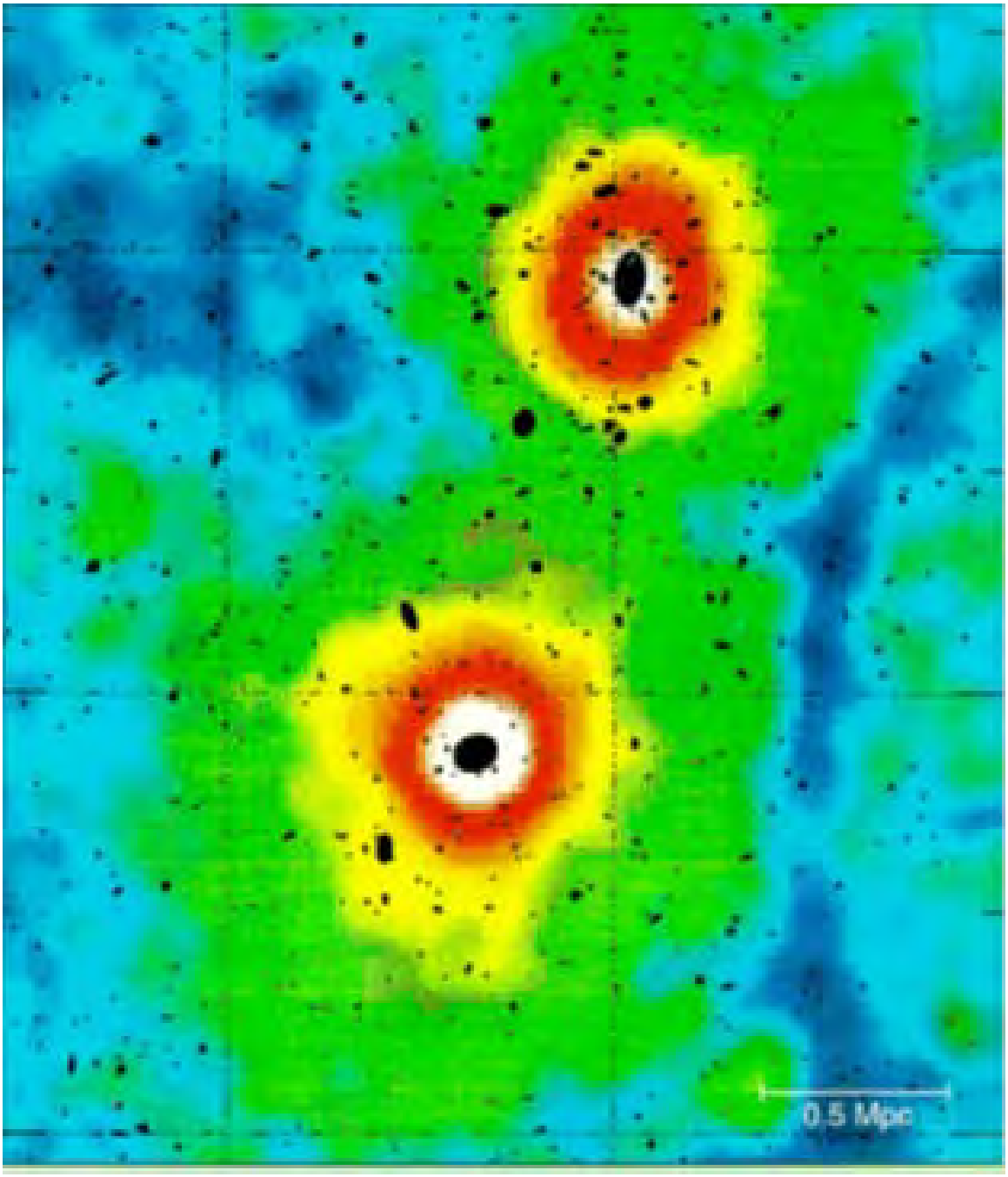
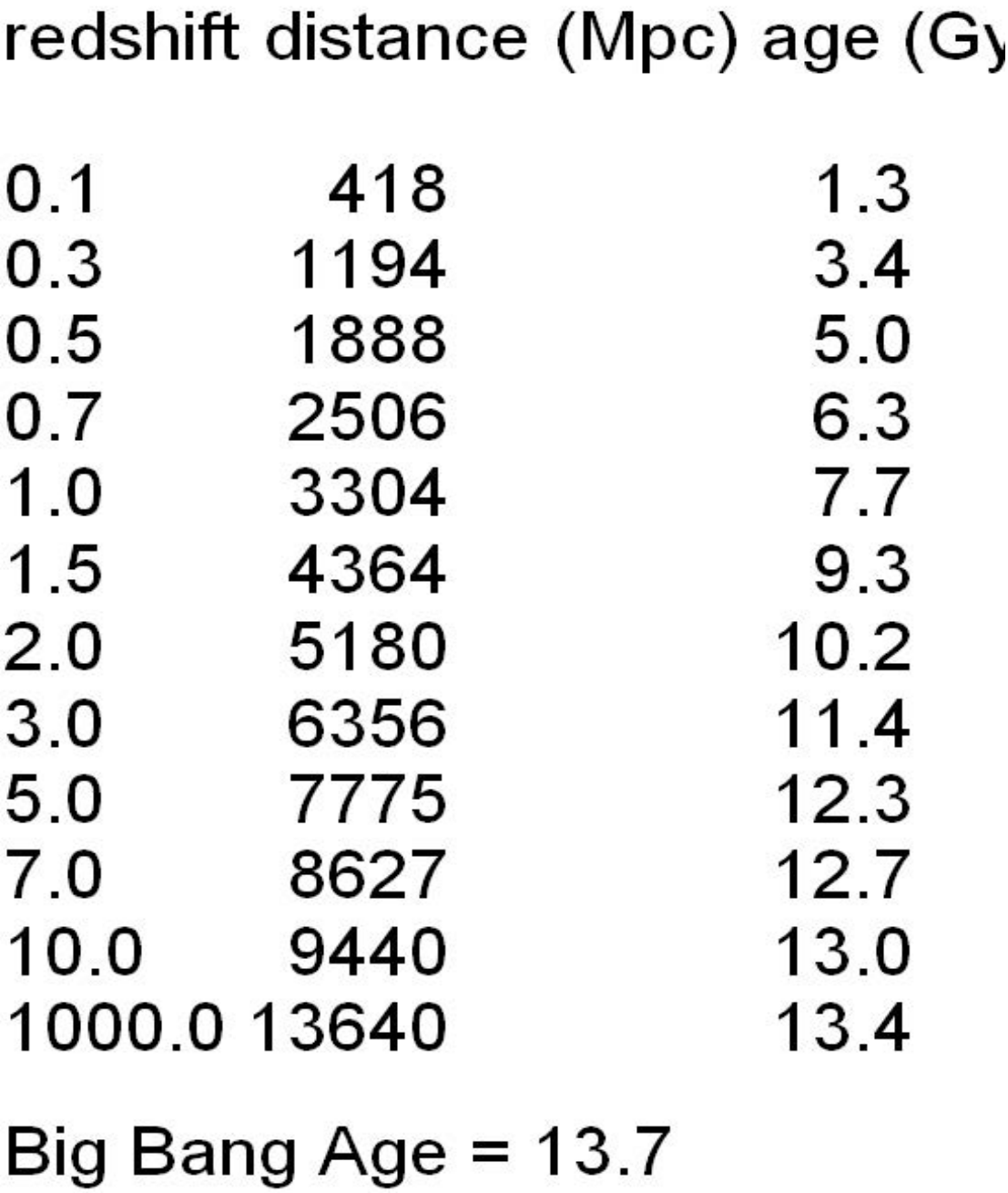


Figure 7: Lookback Timescales



Description and results of a robust approach to $f(R)$ gravity

Luisa G. Jaime[1,2], Leonardo Patiño[2] and Marcelo Salgado[1]

August 30, 2011

Abstract

This contribution is intended to describe a robust method to study the gravitational equations coming from an $f(R)$ theory and give some particular examples by finding spherical gravitational configurations whose existence had been the source of controversy in recent years. We argue that such controversy is the result of an ill defined procedure to treat $f(R)$ theories.

1 Introduction

All the dynamics of general relativity can be extracted from the action

$$S_{EH+matt}[g_{ab}, \boldsymbol{\psi}] = \int \frac{R}{2\kappa} \sqrt{-g}\, d^4x + S_{\rm matt}[g_{ab}, \boldsymbol{\psi}] \ , \tag{1}$$

where the first term on the right is the Einstein-Hilbert action and the second is the one coming from the matter content on the space-time with $\boldsymbol{\psi}$ representing generically the matter fields, $\kappa = 8\pi G$ and we use units where C=1.

Einstein's equation $G_{ab} = \kappa T_{ab}$ can be obtained by varying (1) with respect to the metric g_{ab}.

In the same spirit, $f(R)$ theories are gravitational theories that propose a general, and in principle arbitrary, function f of the curvature scalar R as the gravitational Lagrangian density, modifying (1) to be

$$S[g_{ab}, \boldsymbol{\psi}] = \int \frac{f(R)}{2\kappa} \sqrt{-g}\, d^4x + S_{\rm matt}[g_{ab}, \boldsymbol{\psi}] \ , \tag{2}$$

and from which, by varying it with respect to the metric, we get

$$f_R R_{ab} - \frac{1}{2} f g_{ab} - (\nabla_a \nabla_b - g_{ab}\Box)\, f_R = \kappa T_{ab} \ . \tag{3}$$

where $\Box = g^{ab}\nabla_a \nabla_b$, and $f_R \equiv \frac{df}{dR}$.

The hope behind postulating (2) is to find a gravitational theory that, without introducing dark components of matter and energy, accommodates some of the

Figure 1: Plot of the potentials in the scalar-tensor theory associated with the models: a) In Blue, suggested by Starobinsky, $f(R) = R + \lambda R_0 \left[\left(1 + \frac{R^2}{R_0^2} \right)^{-n} - 1 \right]$, with $\lambda = 2, R_0 = 1$ and $n = 2$, b) In Green, suggested by Miranda *et al*, $f(R) = R - \alpha R_* \ln(1 + \frac{R}{R_*})$, with $\alpha = 2$ and $R_* = 1$, c) In Red, suggested by Hu and Sawicki, $f(R) = -m^2 \frac{c1(\frac{R}{m^2})^n}{c2(\frac{R}{m^2})^n - 1}$, with $m^2 = 1, n = 2, c1 = 1$ and $c2 = 2$.

observations that are not in agreement with general relativity and its weak field limit, such as the rotation curves of the galaxies or the accelerated expansion of the universe.

It should be clarified that the intention of this particular contribution is neither to criticize nor defend the general idea of introducing $f(R)$ theories, but instead to present a reliable method to treat these theories.

Equation (3) is in general a fourth order partial differential equation for the metric coefficients which proves to be very difficult to solve, so it is encouraging no to approach it directly but to consider a simpler way to treat it. In particular in 1983 Teyssandier and Tourrenc [1] suggested a method to rewrite (3) for a certain class of $f(R)$'s by noticing that if $f_{RR} \neq 0$, the action (2) could be rewritten as the action for a scalar-tensor theory

$$S[g_{ab}, \boldsymbol{\psi}] = \int \frac{\sqrt{-g}}{2\kappa} [\chi(\varphi) R - V(\varphi)] \, d^4x + S_{\text{matt}}[g_{ab}, \boldsymbol{\psi}] \, , \tag{4}$$

where

$$\chi = f_R(\varphi) \quad \text{and} \quad V(\varphi) = \varphi f_R(\varphi) - f(\varphi).$$

as long as $\varphi = R$. For (4) to represent a scalar-tensor theory for the field χ, the relation $\chi = f_R(\varphi)$ has to be used to write the potential $V(\varphi)$ as a function of χ.

This method was soon extended [2] to general cases of (2) and clearly simplified matters, nonetheless the mapping from an $f(R)$ theory to a scalar-tensor theory can be ill defined, as it turns out to be the case for several $f(R)$'s commonly found in the literature. This can be seen in figure (1) where we plot the potentials in the scalar-tensor theories that would correspond to the respective $f(R)$ indicated in the caption of the plot. We see that to the $f(R)$ suggested in [3] by Starobinsky corresponds a multivalued potential in the scalar-tensor theory, and the same happens for the $f(R)$ put forward in [4] by Hu and Sawicki. As a consequence of this behavior, the physics extracted from the models using their scalar-tensor counterpart cannot be conclusive.

It is clear from the remarks on this section that a novel way to consistently solve (3) is still necessary, so in the next section we outline one such method which was already discussed in more detail in a previous paper [5].

2 Robust approach to $f(R)$ gravity

It is straightforward to write (3) in the following way

$$f_R G_{ab} - f_{RR}\nabla_a\nabla_b R - f_{RRR}(\nabla_a R)(\nabla_b R) \\ +g_{ab}\left[\frac{1}{2}\left(Rf_R - f\right) + f_{RR}\Box R + f_{RRR}(\nabla R)^2\right] = \kappa T_{ab}\ , \tag{5}$$

where $(\nabla R)^2 := g^{ab}(\nabla_a R)(\nabla_b R)$. Taking the trace of this equation yields

$$\Box R = \frac{1}{3f_{RR}}\left[\kappa T - 3f_{RRR}(\nabla R)^2 + 2f - Rf_R\right]\ , \tag{6}$$

where $T := T^a_a$. Finally, using Eq. (6) in (5) we find

$$G_{ab} = \frac{1}{f_R}\Bigg\{f_{RR}\nabla_a\nabla_b R + f_{RRR}(\nabla_a R)(\nabla_b R) \\ -\frac{g_{ab}}{3}\left[\frac{1}{2}\left(Rf_R + f\right) + \kappa T\right] + \kappa T_{ab}\Bigg\}\ . \tag{7}$$

From the manipulation just carried we see that without any ambiguity, equations (6) and (7) are entirely equivalent to (3), so we propose to use them as the basic equations for $f(R)$ theories in every application.

The equations we intend to use provide second order equations for the metric g_{ab} and the curvature scalar R considered as independent from each other. In the next section we will have the chance to explicitly verify, for spherically symmetric metrics, the advantage of working with (6) and (7) over working with (3) or the equations derived from (4), since (6) and (7) provide a clean and free of pathologies method.

3 Static and spherically symmetric solutions

Let's consider a static and spherically symmetric (SSS) space-time so the metric is given by

$$ds^2 = -n(r)dt^2 + m(r)dr^2 + r^2\left(d\theta^2 + \sin^2\theta d\varphi^2\right), \tag{8}$$

where the metric coefficients n and m are functions of the coordinate r solely.

In this case equation (6) yields,

$$R'' = \frac{1}{3f_{RR}}\Big[m(\kappa T + 2f - Rf_R) - 3f_{RRR}R'^2\Big] + \left(\frac{m'}{2m} - \frac{n'}{2n} - \frac{2}{r}\right)R'\ , \tag{9}$$

where $'$ denotes differentiation with respect to r.

From the $t-t$, $r-r$ and $\theta-\theta$ components of equation (7) and using (9) to abbreviate results, we find

$$\begin{aligned} m' &= \frac{m}{r(2f_R+rR'f_{RR})}\Bigg\{2f_R(1-m)-2mr^2\kappa T^t_{\ t} \\ &+\frac{mr^2}{3}(Rf_R+f+2\kappa T)+\frac{rR'f_{RR}}{f_R}\Big[\frac{mr^2}{3}(2Rf_R-f+\kappa T) \\ &-\kappa mr^2(T^t_{\ t}+T^r_{\ r})+2(1-m)f_R+2rR'f_{RR}\Big]\Bigg\}\,, \end{aligned} \tag{10}$$

$$\begin{aligned} n' &= \frac{n}{r(2f_R+rR'f_{RR})}\Big[mr^2(f-Rf_R+2\kappa T^r_{\ r}) \\ &+2f_R(m-1)-4rR'f_{RR}\Big]\,, \end{aligned} \tag{11}$$

$$\begin{aligned} n'' &= \frac{2nm}{f_R}\Big[\kappa T^\theta_{\ \theta}-\frac{1}{6}(Rf_R+f+2\kappa T)+\frac{R'}{rm}f_{RR}\Big] \\ &+\frac{n}{2r}\Big[2\left(\frac{m'}{m}-\frac{n'}{n}\right)+\frac{rn'}{n}\left(\frac{m'}{m}+\frac{n'}{n}\right)\Big]\,. \end{aligned} \tag{12}$$

At this point we can define first order variables $Q_n = n'$ and $Q_R := R'$ and write equations (9)$-$(12) as $dy^i/dr = \mathcal{F}^i(r, y^i)$ to solve them numerically, where $y^i = (m, n, Q_n, R, Q_R)$. It is important to notice that equations (11) and (12) are not independent, so we can chose one to find the solution we are looking for and the other to perform a consistency check for the numerical method.

We can notice two things about equations (9)$-$(12), one is that if we take $f(R) = R$ they reduce to the equations for SSS coming from general relativity, and the other is that if we consider the expression for R computed directly from (8) in the conventional way,

$$R=\frac{1}{2r^2n^2m^2}\Big[4n^2m(m-1)+rnm'(4n+rn')-2rnm(2n'+rn'')+r^2mn'^2\Big]\,, \tag{13}$$

and substitute n', m' and m'' from (10)-(12) we get the identity $R \equiv R$, confirming the consistency of our equations.

To find a solution we still need information about the matter content, so it is necessary to use the conservation equation $\nabla^a T_{ab} = 0$ satisfied by the matter distribution and to specify an equation of state.

For a perfect fluid with $T_{ab} = (\rho+p)u_a u_b + g_{ab}p$, the conservation equation leads to $p' = -(\rho+p)n'/2n$, which is the modified Tolman-Oppenheimer-Volkoff equation of hydrostatic equilibrium.

In order to solve the differential equations we require some boundary conditions, which in this case are given by the imposition of the desired asymptotic behavior as $r \to \infty$ and regularity at $r = 0$, which will be addressed by expanding the fields around $r = 0$ as

$$\phi(r) = \phi_0 + \phi_2 r^2 + \mathcal{O}(r^4), \tag{14}$$

where ϕ stands for m, n or R.

The odd powers in (14) are set to zero so that the solution is regular at $r = 0$, and the value of ϕ_0 and ϕ_2 will be fixed to impose the desired asymptotic behavior as $r \to \infty$. Without loss of generality, we can impose the local-flatness condition by setting $m_0 = 1$. The metric function $n(r)$ can be normalize to behave as de Sitter, $n(r) \sim 1 - \Lambda_{\text{eff}} r^2/3$, as $r \to \infty$, where Λ_{eff} will be defined below.

We demand that as $r \to \infty$, R goes to a constant R_1 that turns out to be a value for which the function $2f - f_R R$ vanishes, and consequently we decided to introduce a "potential" defined as

$$V(R) = -Rf(R)/3 + \int^R f(x)dx \tag{15}$$

to characterize the possible asymptotic values of R as the critical points of it.

Notice that a constant R_1 satisfying $V'(R_1) \equiv \frac{1}{3}(2f(R) - f_R(R)R) \mid_{R=R_1} = 0$ solves equation (9) exactly in the vacuum region, so that $\Lambda_{\text{eff}} \equiv \frac{1}{4}R_1$ plays the role of a cosmological constant.

From expression (15) we can see that our potential is as well defined as $f(R)$ it self, so we can use it confidently in our analysis. The introduction of (15) proves to be useful when performing numerical explorations, as the case is regularly that R_1 is not only a critical point of it, but actually a minimum.

3.1 Some examples

A particular topic of interest in any $f(R)$ theory concerns the existence of solutions to its gravitational equations that are SSS and represent compact objects. Some debate around this point can be found in the literature, where for an instance regarding the $f(R)$ suggested by Starobinsky it was claimed in [7] that if an incompressible fluid was considered, the solution could not exist because of the appearance of singularities within the object. This assertion was contradicted in [8, 9]. Towards the end of this section we will use our method to find and present finite solutions for incompressible fluids.

For the moment, from the $f(R)$'s in the three boxes of figure (1) we will start with the study of

$$f(R) = R - \alpha R_* \ln\left(1 + R/R_*\right), \tag{16}$$

where α and R_* are positive constants, the latter setting a scale parameter. This $f(R)$ was introduced and studied in [6], where it was mapped to its scalar-tensor counterpart, which as we saw in figure (1) had a well defined potential and we should not expect to disagree with the results found by that method.

In our approach (16) leads to a potential

$$V(\tilde{R}) = \frac{R_*^2}{6}\Big\{(1+\tilde{R})\left[\tilde{R} + (6\alpha - 1)\right] - 2\alpha(3 + 2\tilde{R})\ln\left[1 + \tilde{R}\right]\Big\}, \tag{17}$$

where $\tilde{R} = R/R_*$.

A plot of (17) is given in figure (2) where we can see that it has several critical

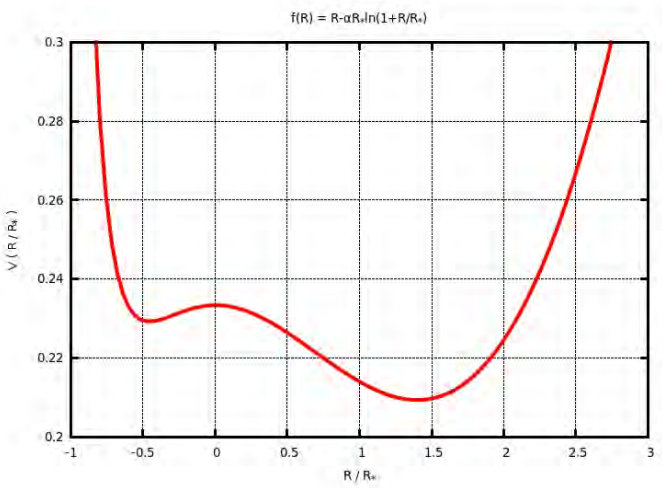


Figure 2: Effective potential associated with $f(R) = R - \alpha R_* \ln(1 + R/R_*)$ with $\alpha = 1.2$

points. To solve the system of equations (9)−(11) we employed a fourth order Runge Kutta algorithm and by a shooting method, varying the value of $\tilde{R}$ at $r = 0$, we found a solution for which as $r \to \infty$, $\tilde{R}$ approaches the global minimum $\tilde{R}_1$ of (17).

Giving the asymptotic behavior we demanded from $n(r)$, $\tilde{R}_1$ corresponds to the de Sitter value and gives rise to an effective cosmological constant $\Lambda_{\text{eff}} = R_1/4$.

The solutions for m, n and R are plotted in figures (3) and (4), where we can see that, as claimed in [6], no singularities are found. We verified that our method reproduced correctly the details of the results found in [6]. To confirm that the numerical method was working properly, we corroborated that all the results we found using the system of equations (9)−(11) were recovered when using (9),(10) and (12) in a separate code.

Now we turn our attention to the $f(R)$ put forward by Starobinsky [3]

$$f(R) = R - \lambda R_* \left\{ 1 - \left[1 + (R/R_*)^2 \right]^{-\beta} \right\} \tag{18}$$

with $\beta = 1$ which was analyzed in [7, 8, 9] using the transformation to a scalar tensor theory. As we saw in figure (1) the potential for (18) is not single valued so it would be expectable to find some discrepancies with the results found previously for this $f(R)$.

In contrast, following our method, the potential associated with (18) is

$$V(\tilde{R}) = \frac{R_*^2}{3} \left\{ \frac{\tilde{R}}{2} \left[\tilde{R} - 4\lambda - 2\lambda \left(1 + \tilde{R}^2 \right)^{-1} \right] - 3\lambda \arctan\left(\tilde{R} \right) \right\}, \tag{19}$$

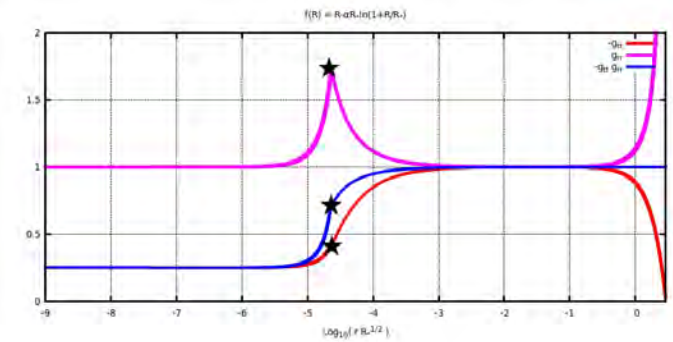


Figure 3: Plot of $n(r) = -g_{tt}$ in red, $m(r) = g_{rr}$ in pink and $n(r)m(r) = -g_{tt}g_{rr}$ in blue for the $f(R)$ suggested by Miranda et al. (16) with $\alpha = 1.2$

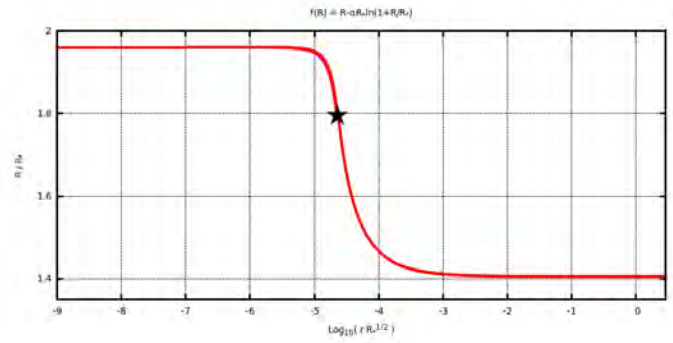


Figure 4: Solution for the Ricci scalar R for the $f(R)$ suggested by Miranda et al. (16) with $\alpha = 1.2$

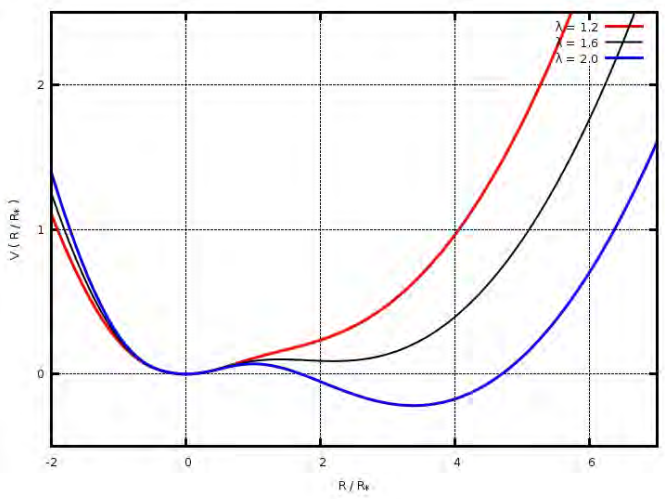


Figure 5: Effective potential associated with the $f(R) = R - \lambda R_* \left\{1 - [1 + (R/R_*)^2]^{-\beta}\right\}$ with $\lambda = 1.2, 1.6$ and 2

which we plot in figure (5) for three values of λ to show that depending on it being below or above $\lambda_{\text{crit}} \sim 1.54$, the potential has respectively one or three critical points.

By carrying the numerical integration in this case and using different values for λ we found several types of solutions like the ones we show in figures 6 and 7. For some values of λ we even found solutions that presented oscillations around $R = 0$ in the limit $r \to \infty$, which we display in figure (8) and that will be analyzed thoroughly elsewhere. The relevant fact for the present discussion is that no singularities were found for R in any of them.

We hope this closes the debate around the existence, or lack thereof, of spherical solution for this theory that was established, as we already mentioned, when in [7] it was claimed that this objects unavoidably developed singularities in the Ricci scalar while in [8, 9] it was assured they didn't. Of course we blame this contradictory results on the ill nature of the potential in the associated scalar-tensor theory that was used to perform the previous studies.

4 Closing remarks

In this contribution we presented a simple way, free of pathologies, to study $f(R)$ theories and introduce the definition of a potential that proved to be a very useful tool.

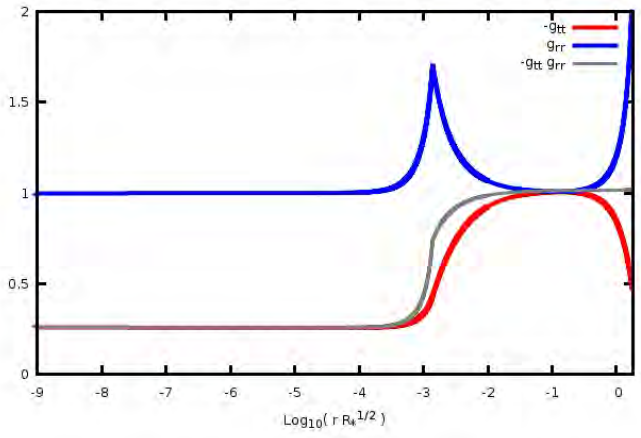


Figure 6: Plot of $n(r) = -g_{tt}$ in red, $m(r) = g_{rr}$ in blue and $n(r)m(r) = -g_{tt}g_{rr}$ in gray for the Starobinsky's model (18) with $\lambda = 1.56$.

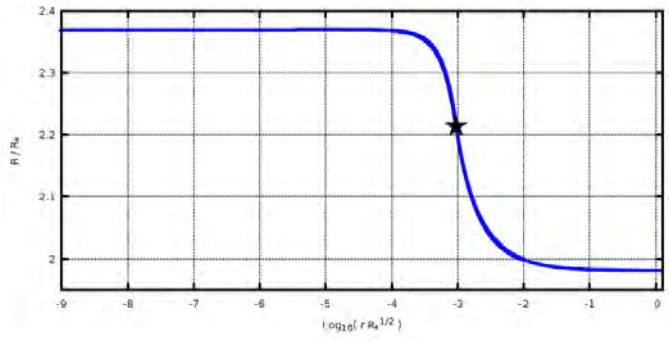


Figure 7: Plot for the Ricci scalar R for the Starobinsky's model (18) with $\lambda = 1.56$.

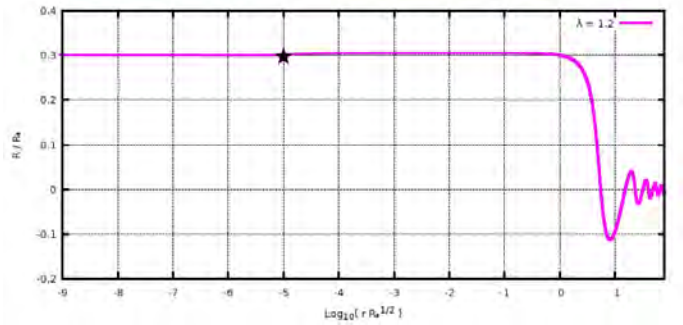


Figure 8: Similar to Fig.7 with $\lambda = 1.2$.

Using some $f(R)$ theories we constructed solutions that had been previously studied in the literature using their mapping to a scalar tensor theory, and found that our results coincide for the cases where the mapping is well defined while making us skeptical about previous results when such a mapping is ill defined. Regarding the existence of certain spherically symmetric solutions, we solved the debate that had been held up to now based on methods that were inconclusive given the lack of mathematical soundness of the correspondence established with scalar tensor theories.

Something that is important to mention is that there is an inherent contrast problem when numerically constructing solutions that should represent realistic astronomical objects. This problem arises since this type of gravitational solutions unavoidably carry two highly dissimilar scales, one corresponding to the density inside the object and the other to the density very far from it. The ratio of these two scales can be of 40 orders of magnitude, hence developing a method that can handle this contrast is still a challenge, and there is work in progress in that direction.

Another scenario that we are currently perusing is the analysis of cosmological solutions, since clearly if there was to be an $f(R)$ theory to describe gravity, it should accommodate all the gravitational configurations that we do observe in nature.

5 Acknowledgments

We would like to thank the organizers of the conference "Two Cosmological Models" for the invitation to present this talk. The research carried for the material presented

here was supported by CONACyT grant SEP-2004-C01-47209-F and by DGAPA-UNAM grants IN119309-3, IN115310 and IN-108309-3.

References

[1] P. Teyssandier and Ph. Tourrenc, J. Math. Phys. **24**, 2793 (1983).

[2] S. Cecotti, Phys. Lett. B **190**, 86 (1987); J. D. Barrow and S. Cotsakis, Phys. Lett. B **214**, 515 (1988); D. Wands, Class. Quant. Grav. **11**, 269 (1994).

[3] A. A. Starobinsky, JETP Lett. **86**, 157 (2007)

[4] W. Hu and I. Sawicki, Phys. Rev. D **76**, 064004 (2007).

[5] L. G. Jaime, L. Patino and M. Salgado, Phys. Rev. D **83**, 024039 (2011).

[6] V. Miranda *et al.*, Phys. Rev. Letters **102**, 221101 (2009)

[7] T. Kobayashi, and K. Maeda, Phys. Rev. D **78**, 064019 (2008)

[8] E. Babichev, and D. Langlois, Phys. Rev. D **80**, 121501(R) (2009)

[9] A. Upadhye, and W. Hu, Phys. Rev. D **80**, 064002 (2009)

General relativistic dynamics of galaxy clusters

F. I. Cooperstock[1]* and S. Tieu[2]

*Department of Physics and Astronomy, University of Victoria,
P.O. Box 3055, Victoria, B.C. V8W 3P6 (Canada)

Abstract. To determine whether there is scope for general relativity to account for the motions of galaxies in clusters without invoking dark matter, we analyze the gravitational collapse of a spherically symmetric ball of dust in the general relativistic weak gravity regime. Using the well-known result that distant external observers do not witness the same motion characteristics as do local observers, we make the velocity comparison for this model. The velocity observed by external observers on the basis of general relativity is derived for the critically open case and is seen to differ markedly from the expectations based upon Newtonian gravity theory. Seen as an idealized model for a cluster of galaxies, we find that with the general relativistic velocity expression, the higher-than-expected constituent velocities observed can be readily correlated with the solely baryonic measure of the mass, removing the need to introduce extraneous dark matter.

In our first paper at this conference, we showed that general relativity could account for the flat galactic rotation curves, the observation of the essentially constant velocities of the stars in the galaxies out to their extremities, without the requirement for the conventionally demanded vast reservoirs of exotic dark matter. The central departure from conventional thinking revolves around the fact that general relativistic nonlinearities can play an important role even in the context of weak gravity. While there was some independent support for our work [3], there were several papers offering an interesting variety of critical remarks. It should be stressed that we have responded to *all* of our critics and we stand behind our work. It should also be stressed however, that while we can account for the high velocity motions of the stars without invoking vast reservoirs of dark matter in galactic halos, this is based on presently available data. We described a velocity dispersion test using data of stellar motions above and below the galactic symmetry plane to determine the extent of extra matter that may exist beyond the visible matter.

The single galaxy work that we had described was based on a stationary (i.e. no explicit time-dependence) solution of the Einstein field equations and it was an approximate solution. As well, it contained a benign singularity, a density gradient discontinuity, which was incorrectly interpreted by our critics as a manifestation of a negative mass surface layer equal in magnitude to the entire mass of the galaxy. We demonstrated that this was an improper interpretation but as we know, critics are not, as a rule, readily won over.

In searching for a new example to test the effects of general relativity on a collective

[1] cooperst@uvic.ca
[2] stieu@uvic.ca

of freely gravitating matter, we seized upon the *exact* solution of spherical gravitational collapse of a ball of dust [4] (Further details of the present work can be found in [4].) The advantages in this are evident: a) with the solution being exact, approximation issues do not arise b) the solution is dynamic, adding a new interesting aspect and c) the solution is totally singularity-free in the phase of analysis under consideration. Specifically, the analysis is for the phase in which the last dust elements have already reached the point of forming a discernable ball of material (with purely vacuum beyond the outer surface) and long before a concentration leading to singularity formation can occur. While the interest in the collapse problem in the past has focused upon strong gravity leading to singularity formation (see, e.g. [5] as well as the pioneering paper on gravitational collapse [6]), we emphasize that we consider here the *weak* gravity regime, long before any singularity could be formed. Even though highly idealized, this could be regarded as a special case model of a cluster of galaxies in evolution where, in the normally unsymmetric case, the component velocities have been observed to have unusually high velocities, that is to say high compared to the expectation of Newtonian gravity but still very much smaller than the speed of light. According to standard accounts, this phenomenon was the historic origin of the dark matter hypothesis, advanced in the 1930's by Zwicky, in an effort to explain the high velocity observations within the context of Newtonian gravity theory. In this paper, we show that in this idealized model, such relatively high velocities can be accounted for in principle using general relativity in the absence of an additional contribution from dark matter.

It is important to distinguish between the general relativistic effects of test particles falling *in vacuum* towards a concentrated central mass versus particles falling freely *as a collective* in the case of dust with the elements gravitationally interacting with each other. To do so, we first consider the treatment in the classic Landau-Lifshitz text [7] (henceforth referred to as "LL") of the familiar Schwarzschild solution, the spherically symmetric vacuum gravitational field generated by a spherically symmetric central mass m. A spherically symmetric metric can be expressed in generality in spherical polar coordinates in the form

$$ds^2 = e^{\nu(r,t)}dt^2 - e^{\Lambda(r,t)}dr^2 - r^2(d\theta^2 + \sin^2\theta d\varphi^2) \tag{1}$$

with units chosen with $c = G = 1$.

In the case of vacuum, the metric functions are readily found by solving the Einstein field equations and reflecting the intrinsically static nature of the spherically symmetric vacuum solution (Birkhoff's theorem), can be expressed in the time-independent form

$$e^{\nu} = 1 - \frac{2m}{r}, \qquad e^{\Lambda} = \left(1 - \frac{2m}{r}\right)^{-1}. \tag{2}$$

the well-known "Schwarzschild metric". In [7], LL transform to comoving "synchronous" coordinates (R, τ) as

$$\begin{aligned} \tau &= t + \int \frac{f(r)}{1 - \frac{2m}{r}}\, dr \\ R &= t + \int \frac{1}{f(r)\left(1 - \frac{2m}{r}\right)}\, dr \end{aligned} \tag{3}$$

while leaving θ and ϕ unchanged. In these coordinates, an observer who is at $R =$ *constant* is physically in the state of free-fall.

With $f(r)$ chosen as

$$f(r) = \sqrt{\frac{2m}{r}} \tag{4}$$

we find the simple relationship between the coordinates

$$r = \left(\frac{3}{2}(R-\tau)\right)^{2/3}(2m)^{1/3}. \tag{5}$$

By so doing, the singularity issues encountered in (1),(2) at $r = 2m$ are removed. The metric in these new comoving (R,τ) coordinates is then expressed as

$$ds^2 = d\tau^2 - \frac{dR^2}{\left(\frac{3}{2(2m)}(R-\tau)\right)^{2/3}} - \left(\frac{3}{2}(R-\tau)\right)^{4/3}(2m)^{2/3}(d\theta^2+\sin^2\theta d\varphi^2) \tag{6}$$

which is now seen to be time-dependent (naturally so, as a free-fall observer sees a steadily changing picture of the central mass drawing ever closer with the elapse of time). The standard analysis revolves around the strong gravity regime which evolves as the proper time τ approaches R and the singularity emerges at $\tau = R$.

By contrast, our focus will be on the weak gravity regime where $\tau \ll R$ for all R. This translates to $r \gg 2m$ for all r in the (r,t) frame. Up to the time of our work, such a focus would likely have never been contemplated: one would have tended to believe that one might just as well use Newtonian gravity for such weak fields.

The time coordinate t measures time read by the asymptotic observer. The standard general relativistic treatment concentrates upon the regime of strong gravity where the difference in perception of the proper velocity of a freely falling test particle as measured by the local observer in comparison to the measurement of the velocity by the asymptotic observer becomes particularly significant (see [7]).

In generality the proper radial velocity of a freely falling test particle cannot be evaluated in the (R,τ) coordinates. This is because R is constant for any given particle in this frame and therefore the radial velocity is always zero in this comoving frame. For the required ingredients, LL use the solution of the radial geodesic equation for a freely falling test particle in the usual (non-comoving) Schwarzschild coordinates (r,t) as employed in (1),(2) which are suitable for this purpose. Using the notation $(x^0,x^1,x^2,x^3) \equiv (t,r,\theta,\phi)$, the geodesic solution for dr/dt and the metric coefficients g_{00} and g_{11} of (1) are used to evaluate the proper radial velocity

$$v = -\sqrt{\frac{-g_{11}}{g_{00}}}\frac{dr}{dt}. \tag{7}$$

This equals $\sqrt{2m/r}$ in magnitude for particles released from rest at infinity and is seen to approach 1, the speed of light, as r approaches $2m$. (The rest release point r_0 in LL Eq.(102.7) is taken to be infinite here.) However, for asymptotic observers who reckon

radial distance and time increments as dr and dt, the measured velocity is

$$\frac{dr}{dt} = -\left(1-\frac{2m}{r}\right)\sqrt{\frac{2m}{r}} \tag{8}$$

which approaches zero in the very strong gravity regime as r approaches $2m$. This is in stark contrast to the proper radial velocity. However, for weak gravity which is our focus in this paper, the $(1-2m/r)$ factor in (8) is approximately 1 and the local proper and asymptotic measures of velocity are approximately equal in the value $-\sqrt{2m/r}$. This justifies the neglect of general relativity in the context of weak gravity for the case of test particle motion in spherically symmetric vacuum. From this, we can appreciate the bias that investigators would have regarding motions of particles and their observations for the case of weak fields. However motion of a single test particle is not the same as the motion of a particle that is part of a collective, gravitationally interacting with all of the particles. The collective has a life of its own in general relativity as we will witness when we turn to the analysis of spherically symmetric dust collapse.

As with the vacuum case, LL choose comoving coordinates for dust collapse and structurally as in (6), express the metric as

$$ds^2 = d\tau^2 - e^{\lambda(\tau,R)}dR^2 - r^2(\tau,R)(d\theta^2+\sin^2\theta d\varphi^2). \tag{9}$$

The four non-vanishing Einstein field equations are extremely non-linear:

$$-e^{-\lambda}(r')^2 + 2r\ddot{r} + \dot{r}^2 + 1 = 0, \tag{10}$$

$$-\frac{e^{-\lambda}}{r}\left(2r'' - r'\lambda'\right) + \frac{\dot{r}\dot{\lambda}}{r} + \ddot{\lambda} + \frac{\dot{\lambda}^2}{2} + \frac{2\ddot{r}}{r} = 0 \tag{11}$$

$$-\frac{e^{-\lambda}}{r^2}\left(2rr'' + (r')^2 - rr'\lambda'\right) + \frac{1}{r^2}\left(r\dot{r}\dot{\lambda} + \dot{r}^2 + 1\right) = 8\pi\rho \tag{12}$$

$$2(\dot{r})' - \dot{\lambda}r' = 0. \tag{13}$$

In the above, a dot denotes the partial derivative with respect to τ and a prime denotes the partial derivative with respect to R.

It is remarkable that for such a complicated non-linear set of partial differential equations, the solution should be expressible in the simple form:

$$e^{\lambda} = \frac{(r')^2}{1+E(R)} \tag{14}$$

$$\dot{r}^2 = E(R) + \frac{F(R)}{r} \tag{15}$$

$$r = \left(\frac{9F}{4}\right)^{1/3}(R-\tau)^{2/3} \tag{16}$$

($E(R)$ and $F(R)$ are functions of integration) where in (16), we have chosen the part of the three-stage solution [7] for the case where $E(R)$ in (15) is taken to be zero, the

particles being released from rest at infinity in the infinitely distant past (the critically open case). The simplicity of appearance is deceptive however, since it is only with a simultaneous use of both coordinate systems that the solution has a simple appearance. In fact the appearance could lead one to believe that the solution is really the same as is the case in Newtonian gravity. This would be a serious error as we shall witness in what follows.

There are two other cases for non-zero E, the bound and the unbound cases, which are familiar in concept from classical mechanics. These solutions are expressed in parametrized form in [7]. In the present study, we focus upon the simplest critically open case $E = 0$. For all three cases, positive, negative or zero E, the density ρ is incorporated into the solution as

$$8\pi\rho = \frac{F'}{r'r^2}. \tag{17}$$

From (17), a simple integration (see [7]) shows that the mass $M(R)$ within the radial coordinate R is

$$M(R) = F(R)/2 \tag{18}$$

and thus the entire mass M is given by $M(R_0)$ where R_0 is the outer comoving radial coordinate of the dust ball.

Our focus here is upon the radial dust velocity measured by distant "rest" (i.e. *non*-comoving) observers. As with the vacuum case, we must choose new coordinates for the evaluation of the radial velocity because the R coordinate is constant for any given dust particle. For the dust case, we continue to use the approach taken by LL for vacuum and evaluate this radial velocity dr/dt in the Schwarzschild-like (r,t) coordinate frame. However, unlike the case of vacuum, it is unnecessary to solve the geodesic equations at this point because the (geodesic pressure-free) motion of the dust medium has already been solved in the comoving frame. It is this motion that is of concern to us. What is required is to re-express the solution in Schwarzschild-like (r,t) coordinates.

For consistency with the solution form of (16) and to maintain maximum available generality, we choose the general form of transformation with arbitrary functions $p(r,t)$ and $q(r,t)$,

$$\sqrt{F}R = p(r,t) \tag{19}$$

$$\sqrt{F}\tau = q(r,t) \tag{20}$$

with the constraint

$$p(r,t) - q(r,t) = (2/3)r^{3/2}. \tag{21}$$

From (21), we see that

$$p'(r,t) - q'(r,t) = r^{1/2}, \dot{p}(r,t) = \dot{q}(r,t) \tag{22}$$

where an over-dot denotes $\partial/\partial t$ and a prime denotes $\partial/\partial r$ *only* when acting upon p and q. Note that these derivative symbols elsewhere refer to partial derivatives with respect to τ and R respectively.

We take differentials of (19), (20) and with (21),(22), solve for dR and $d\tau$. These differentials are substituted into (9) to derive the normal form of the metric in Schwarzschild-like coordinates (r,t) with terms of the form $g_{00}dt^2$ and $g_{rr}dr^2$, as well as an undesired cross-term of the form $2g_{0r}drdt$. This cross-term must vanish to mesh with the exterior Schwarzschild metric at the vacuum interface and maintain the useful Schwarzschild-like diagonal form within the ball. This metric form includes a yet-to-be-determined $p'(r,t)$ which we set to diagonalize the metric by making $g_{0r}=0$ yielding

$$p' = \frac{(\frac{3R\sqrt{F}\alpha}{2r} + \sqrt{r}\beta)}{(\alpha+\beta)(1-\beta^2)} \tag{23}$$

where

$$\begin{aligned} \alpha &= \frac{rF'}{3F} = \frac{rM'(R)}{3M(R)} \\ \beta &= \sqrt{\frac{F}{r}} = \sqrt{\frac{2M(R)}{r}}. \end{aligned} \tag{24}$$

Also required in the calculation for p' is e^{λ} which, from (14), is equal to $(r')^2$ for $E=0$. In turn, this requires r' which is computed from (16) yielding

$$r' = \alpha+\beta. \tag{25}$$

Since the R coordinate is comoving with the matter, we express the condition for the radial motion of the particles by taking differentials of (19) and setting $dR=0$ (for motion with spherical symmetry, $d\theta$=$d\phi$=0):

$$p'(r,t)dr + \dot{p}(r,t)dt = 0 \tag{26}$$

from which we find the form of the radial velocity of the particles as judged by external observers

$$dr/dt = -\dot{p}(r,t)/p'(r,t). \tag{27}$$

To solve for $\dot{p}(r,t)$, we first apply $\partial/\partial t$ to (17):

$$8\pi\frac{\partial\rho}{\partial t} = \frac{F'^2\left(\frac{\alpha}{F}+\beta(\frac{F''}{F'^2}-\frac{1}{2F})\right)\dot{p}}{r^2(\alpha+\beta)^2(\frac{3R\sqrt{F}\alpha}{2r}+\sqrt{r}\beta)}. \tag{28}$$

The derivation of (28) made use of (25)

$$\frac{\partial(\alpha+\beta)}{\partial t} = \left[\frac{F'}{2\sqrt{Fr}} + \frac{r}{3}\left(\frac{F''}{F} - \frac{F'^2}{F^2}\right)\right]\frac{\partial R}{\partial t} \tag{29}$$

(where (24) has been used) and the elimination of $\frac{\partial R}{\partial t}$ using

$$\dot{p} = \left[\frac{F'R}{2\sqrt{F}} + \sqrt{F}\right]\frac{\partial R}{\partial t} \tag{30}$$

which follows from the partial differentiation with respect to t of (19). Finally, using (23) and (28) in conjunction with (27) (and with a cancellation of the factor $(\frac{3R\sqrt{F}\alpha}{2r} + \sqrt{r}\beta)$), we find

$$\frac{dr}{dt} = -\frac{(\alpha+\beta)(1-\beta^2)}{8\pi r^2\rho^2}\left[\frac{\alpha}{F} + \beta\left(\frac{F''}{(F')^2} - \frac{1}{2F}\right)\right]^{-1}\frac{\partial\rho}{\partial t}. \tag{31}$$

This is the key equation. The complexity of this velocity expression as computed by observers external to the distribution of matter is in very sharp contrast to the simplicity of the proper velocity form $\beta = \sqrt{\frac{F}{r}}$ as witnessed by local observers. However, *it is* $\frac{dr}{dt}$ *that is the required quantity for astronomical observers.* The velocity is deduced from the Doppler red shift of spectral lines. This shift arises from the relative velocity of emitter to absorber which is dr/dt as judged by distant observers. Thus, dr/dt is to be used in the Doppler formula. The local proper velocity would only be used if the observers were within the vicinity of the emitting galaxies, immersed within the collective swarm. There is also a gravitational red shift but this is generally of higher order for weak fields and small velocities.

By contrast, for local observers, it is the mass $M(R) = F(R)/2$ at radii within the point of interest, that determines the velocity. This is the same as the situation in Newtonian gravity where for spherical symmetry, no other factors such as local density can affect the velocity. However, we see in (31) that in general relativity, the external observers base their perceived velocity measurements on additional factors, the reciprocal of the local density squared and its time rate of change (also expressible as the time rate of change of reciprocal density), the gradient of the mass within the radius in question, $M'(R)$ as well as well as its gradient, $M''(R)$.

It is interesting to note that in the limit of very strong gravity, with β approaching 1, the situation is the same as we witnessed in vacuum: the local observers see the velocity approach 1 whereas the external observers see the velocity approach 0.

It certainly comes as a surprise that for weak gravity, with $\beta \ll 1$, the vacuum and dust comparison is very different. While the measurements for local and asymptotic velocity for observers plotting freely falling test particles in vacuum in the field of a concentrated mass are approximately the same, namely β, the corresponding velocities for local and asymptotic measurements for dust are very different in general: the velocity is simply β for the local measurement whereas the asymptotic measurement is given by the rich expression (31) with $1-\beta^2$ approximated by 1. Indeed, given that the form of dr/dt in (31) is so complicated, it would be a very special occurrence for dr/dt to have the value β. Thus, when astronomers were surprised to witness velocities greater than β in galactic clusters, with what we now know, it would have been more unusual had they witnessed precisely β velocities as there are the various other factors that go into the net velocity expression.

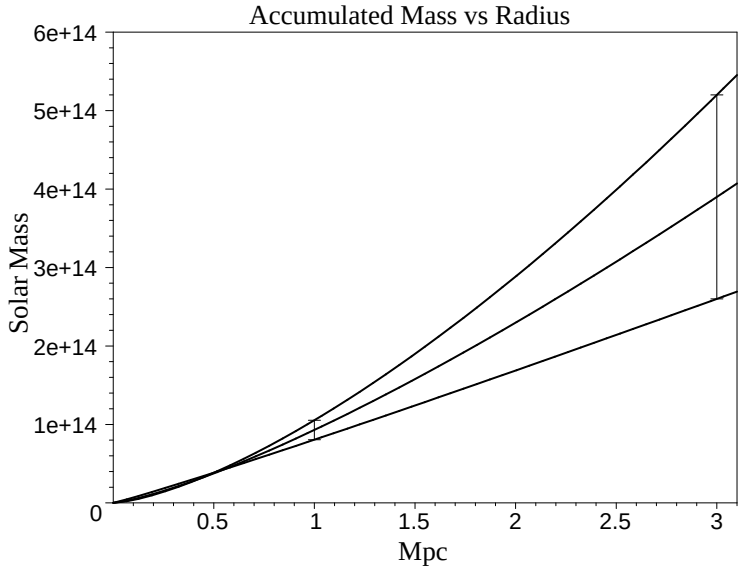


FIGURE 1. The upper, middle and lower limits of mass accumulation are described by the functions, $F = 6.641 \times 10^{-16} R^{1.453}$, $F = 1.244 \times 10^{-12} R^{1.305}$ and $F = 2.531 \times 10^{-7} R^{1.066}$ respectively.

It is of considerable interest to apply (31) to the astrophysical realm. At most locations in the universe, the gravitational field is weak and Newtonian theory has been deemed to be perfectly adequate. Thus, when galaxies in clusters, with gravity found to be weak, were seen to have velocities exceeding β, dark matter was introduced as the necessary mass booster to align the observations with enlarged β. Newtonian theory formed the basis for the calculations. However, we have seen that general relativity, which is essentially universally accepted as the preferred theory of gravity, actually predicts velocities that have elements beyond β even when the gravity is weak. *An essential point is that the nonlinearities of general relativity involving the interactions between the elements under investigation, play an important role in this problem, leading to expressly non-Newtonian behaviour, even though the gravity is weak.*

For the Coma Cluster of galaxies, the ratio $2M(R_0)/r_0$ is of order 10^{-4} assuming the existence of dark matter and of the order 10^{-5} by not assuming any dark matter. The gravity is indeed very weak in this source for the kind of applications to the various galactic clusters under consideration. Thus as a test model, we consider an idealized Coma Cluster of galaxies, one of spherical symmetry with the velocities as reported in [8]. At a radius of 1 Mpc, the total cluster mass, including dark matter, is given as $6.2 \times 10^{14} M_\odot$, with the 13%-17% portion being normal baryonic matter. Within a radius of 3 Mpc, the total mass is reported to be $1.3 \times 10^{15} M_\odot$, with the normal luminous matter portion within the wide range of 20%-40%.

We easily fit these data with an accumulated mass function

$$F(R) = k_1 R^{k_2}, \tag{32}$$

(k_1, k_2 constants) as shown in Figure 1. With $F(0) = 0$ from (32), we are assured that there is no singularity at the origin [7]. Using (32) in (17), we derive the density profile

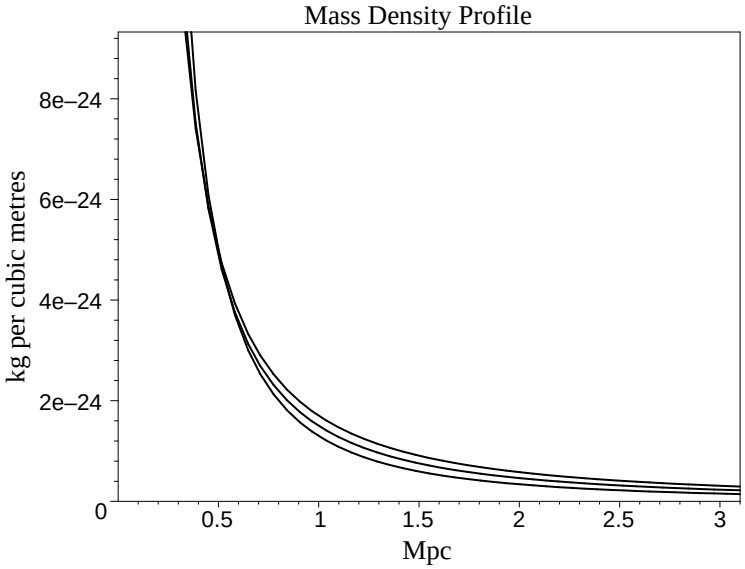


FIGURE 2. From the three functions, $F = 6.641 \times 10^{-16} R^{1.453}$, $F = 1.244 \times 10^{-12} R^{1.305}$ and $F = 2.531 \times 10^{-7} R^{1.066}$, we can derive the mass density as shown in the graph.

for the distribution. The graph of the densities for the two extremes of the uncertainty range and the average is shown in Figure 2.

The velocity associated with each $F(R)$ is given by (31) where we can set the "boosted" velocity as

$$dr/dt = -n\beta \tag{33}$$

where β, as throughout the paper, is composed from the baryonic mass alone and n is the "booster" number to bring dr/dt to the observed higher level of velocity. Assuming the baryonic mass is 20%, 30% and 40% of $1.3 \times 10^{15} M_{\odot}$, we find that the boost factors n are 2.23, 1.82 and 1.58, respectively. Applying this to (31), we can solve for $\partial\rho/\partial t$, which is the only unknown factor. The results are: 2.13×10^{-41}kg/m^3/sec, 2.62×10^{-41}kg/m^3/sec and 3.02×10^{-41}kg/m^3/sec, respectively. Rates such as 10^{-41}kg/m^3/sec are quite reasonable as over a period of one billion years, the density would grow by 10^{-25} kg/m^3, thus roughly doubling the value of the present density.

In this example, we see adequate scope to explain the observed velocities within the framework of general relativity without the requirement of any extraneous dark matter. The new elements of local density, its time rate of change, the gradient of the mass interior to the observation point as well as its gradient are additional factors that ultimately determine the net observed velocity of the matter by external observers. While this is an idealized case of perfect spherical symmetry, it would seem reasonable to expect comparable effects for non-spherical accumulations of freely-gravitating collections of bodies as exist in clusters of galaxies.

It is interesting to consider how the evolution in astronomical thinking might have differed had astronomers applied general relativity rather than Newtonian gravity to galactic dynamics in the early years. It is also significant that astronomers continue to

rely upon Newtonian gravity for their numerical analyses, deploying vast amounts of computer power, time and effort. We feel that it is important to get the message relayed to them that Einstein's gravity is being ignored at their peril.

There is considerable further analysis ahead. To this point we have only dealt with the simplest case $E(R) = 0$. The positive and negative cases for E offer greater freedom of expression. Ultimately, the ideal would be to formulate the equivalent effects of general relativity as applied to *chaotic* weakly- gravitating systems. For this, the general relativistic equivalent of the virial theorem is called for. This is particularly important for the galaxies closer to the centers of clusters as these are more strongly virialized. However, the spherically symmetric treatment in this paper would be more closely connected with the galaxies at the outer regions of clusters. As well, there is the issue of the interpretation of lensing as a mechanism for the deduction of mass. The subtleties of general relativistic weak gravity that we have found in the present work must now be directed to the consideration of lensing. In addition, the temperature of the baryonic inter-galactic gas as seen in Xray emission will have to be taken into account properly.

It is important to note that as before ([1], [2]), we are seeing here the power of the nonlinearities inherent in general relativity in the context of *weak* gravity to produce very significant changes relative to the results expected on the basis of Newtonian theory.

Noting that there are various new elements that come into play in weak field collapse from the vantage point of the external observers, it would be particularly valuable if a laboratory or space-based realization of a spherical collapse could be implemented. If such a design could actually measure with some precision the realization of these elements of dependency, it could serve as a new test of general relativity.

REFERENCES

1. F.I. Cooperstock, and S. Tieu, *Mod. Phys. Lett. A* **21**, 2133, 2006.
2. F.I. Cooperstock and S. Tieu, *Int. J. Mod. Phys. A* **22**, No. 13, 2293, 2007.
3. H. Balasin and D. Grumiller, astro-ph/0602519
4. F.I. Cooperstock, and S. Tieu, *Mod. Phys. Lett. A* **23**, 1745, 2008.
5. F.I. Cooperstock, S. Jhingan, P.S. Joshi and T.P. Singh, *Class. Quantum Grav.* **14**, 2195 (1997).
6. J.R. Oppenheimer and H. Snyder, *Phys. Rev.* **56**, 455 (1939).
7. L.D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* Fourth revised English edition, Pergamon Press, Oxford, 1975.
8. J.P. Hughes, astro-ph/9709272

A Good Fit to the Missing Mass Problem in Galaxies and Clusters of Galaxies

Joel R. Brownstein

University of Utah, Department of Physics & Astronomy, UT 84112

Abstract. Precise measurements of the distribution of matter in galaxies and clusters of galaxies is essential to resolving the missing mass problem. A universal fitting formula for the distribution of dark matter is revealed to fit galaxy rotation curves of a sample of 29 galaxies, including high and low surface brightness and dwarf galaxies in the Ursa Major filament, without introducing the cusp problem of singular halo profiles. The asymptotic density profile is well described by a steep power-law, $\rho(r) \propto r^{-3}$, at large distances from the center of the galaxy, and an increasingly shallow power-law toward the center of the galaxy approaching a rarified, constant density core, where visible baryons dominate the Newtonian force law. The same universal fitting formula is revealed to fit cluster mass profiles of a sample of 10 X-ray clusters of galaxies. The variation of the ratio of the inferred Newtonian dynamic mass to the observed baryon mass is compared across the scales, demonstrating the crucial role played by baryons, and defining what a good theory must achieve in order to provide a good fit to the missing mass problem, and provide a test bed for the predictions of general relativity, constraining modified theories of gravity.

Keywords: Masses and mass distribution, Dark Matter, Modified Theories of Gravity
PACS: 98.62.Ck,95.35.+d,04.50.Kd

1. INTRODUCTION

Enormous strides have been taken to understand the growth of the large scale structure of the Universe from studying the anisotropy power spectrum of the cosmic microwave background [1, 2], and cross-correlating these measurements with the baryon acoustic oscillations imprinted on the spatial distribution of the clustering of galaxies [3, 4, 5], and with the luminosity distances of Type Ia Supernovae [6]. Flat (or nearly-flat) cosmological models with a mixture of radiation, ordinary baryonic matter, cold collisionless dark matter and cosmological constant (or quintessence) and a nearly scale-invariant adiabatic spectrum of primordial density fluctuations provide remarkable fits to large scale ($\gg$ 1 Mpc) observations. While substantial progress has been made in constraining the flat-ΛCDM cosmological parameters to within a few percentages [7], there is still a lack of understanding of the evolution of the baryonic component, which is deeply connected with the formation and evolution of structure at various scales [8]. There remains a large amount of data on galactic and sub-galactic scales ($\ll$ 100 kpc) which require universal explanation [9]. As computing power increases, and hydrodynamical simulations improve, it is unclear whether the inner dark matter density converges to a single power-law form, rather than becoming progressively shallower in slope at smaller radii [10, 11]. Analysis of galaxy rotation curves for late-type spiral galaxies indicate that they must have shallow, non-cusped dark matter profiles, most noticeable for dwarf galaxies [12, 13, 14]. Analysis of X-ray clusters of galaxies indicate inner slopes for the

total mass profiles ranging from steep to shallow, most shallow for the smallest concentration of the gas mass fraction, [14, 15].

Our current knowledge of the baryon component comes from multiwavelength observations, including the 21 cm emission for neutral HI, infrared photometry of the interstellar medium, X-ray spectral imaging for the $> 10^7 K$ intracluster medium, radio synchrotron emission for high-energy plasma, and the velocity and turbulence fields of galaxies and clusters of galaxies. Combined multiwavelength observations culminate in high-resolution galaxy rotation curves and X-ray cluster images [14]. Outstanding scientific questions include:

- How does the Universe evolve from large linear scales dominated by dark matter to the non-linear scales of galaxies and clusters, where baryons and dark matter both play important, interacting, roles?
- Do relaxed structures exhibit steep or increasingly shallow density profiles in the central regions due to cores of dense baryons?
- What is the dark matter self-interaction coupling strength? Can the nature of dark matter be sufficiently constrained to predict a candidate particle from the list of standard model extensions?
- In the absence of dominant dark matter, is there a modification to the theory of gravity which can solve the missing mass problem, at all scales, universally?

To address these questions, the technique of imaging dark matter halos is presented in §2, and we present what is known about the global mass to light ratio in §3 by considering a sample of 19 high surface brightness (HSB) spiral galaxies and 10 low surface brightness (LSB) spiral galaxies from the Ursa Major filament in §3.1 and a sample of 10 X-ray clusters of galaxies in §3.2. Conclusions are drawn in §4.

2. IMAGING DARK MATTER HALOS

Ostriker et al. [16] argued that the masses of ordinary galaxies – found by assuming a constant mass-to-light ratio – may have been underestimated by a factor of 10; but that the galaxy rotation curves in the inner regions provide almost no information about the exterior halo masses. Upon application of a Newtonian force law,

$$a(r) = -\frac{G_N M(r)}{r^2}, \tag{1}$$

where G_N is Newton's constant measured experimentally[1], one may obtain the Newtonian dynamic mass, $M(r)$, which is the mass interior to the sphere of radius, r, needed to support the galaxy rotation curve.

Ostriker et al. [16] observed that although the surface luminosity profiles, $L(r)$, do appear to be convergent, the Newtonian dynamic masses, $M(r)$, diverge with r either

[1] $G_N = 6.67428(67) \times 10^{-11} \mathrm{m}^3 \mathrm{kg}^{-1} \mathrm{s}^{-2}$ [NIST 2006 CODATA value]

weakly (logarithmic) or strongly (linear) depending on the method of measurement, and concluded that within local giant spiral galaxies,

$$M(r) \propto r \qquad \text{for } 20 \text{ kpc } \leq r \leq 500 \text{ kpc}. \tag{2}$$

High-resolution galaxy rotation curves, for both HSB and LSB spiral galaxies, can be fitted to the total mass density profile

$$\rho \propto r^{-\gamma(r)}. \tag{3}$$

The fits grossly reproduce the observed flat rotation curves with $\langle\gamma\rangle = 2$, corresponding to mass profiles which diverge linearly with distance. The divergent mass-to-light ratio necessitates the existence of giant halos surrounding ordinary galaxies of dark matter – the implied density distribution similar to isothermal gas spheres in the outer parts, [17]

$$\rho(r) = \frac{\rho_0 r_c^2}{r^2 + r_c^2}, \tag{4}$$

where r_c is the core radius and ρ_0 is the central dark matter density. In the limit of small $r \ll r_c$, the isothermal sphere model approaches a constant density core. Spherically integrating the constant density core model of Equation (4) one obtains a simple fitting formula for the mass of dark matter,

$$M(r) = 4\pi\rho_0 r_c^3 \left\{ \frac{r}{r_c} - \tan^{-1}(r/r_c) \right\}, \tag{5}$$

which diverges with the behavior of Equation (2), for $r \gg r_c$.

In search of a universal description of collisionless dark matter, Navarro et al. [18, 19] provided power-law fits to halo density profiles using N-body simulations, showing that halo profiles are shallower than r^{-2} near the center and steeper than r^{-2} near the virial radius. The NFW profile is then a simple fitting formula to Equation (3), with a radially *varying* power-law $1 \leq \gamma(r) \leq 3$, to describe spherically averaged density profiles:

$$\rho(r) = \frac{\rho_0 r_s^3}{r(r + r_s)^2}. \tag{6}$$

Spherically integrating the NFW profile of Equation (6) one obtains a simple formula for the mass of dark matter,

$$M(r) = 4\pi\rho_0 r_s^3 \left\{ \ln(r + r_s) - \ln(r_s) - \frac{r}{r + r_s} \right\}, \tag{7}$$

which diverges logarithmically, for $r \gg r_s$. In the limit of small $r \ll r_s$, the NFW fitting formula of Equation (6) approaches the power-law with $\gamma \to 1$; and in the limit of large $r \gg r_s$ approaches the power-law with $\gamma \to 3$ – which does not approximate isothermal spheres. Navarro et al. [18] reported that rotation curves from galaxies ranging in size from giant to dwarf, satellites and gaseous atmospheres are compatible with the NFW

halo structure of Equation (6) provided the mass-to-light ratio increases with luminosity. Navarro et al. [18] determined that the central regions of the NFW distribution have densities comparable to the luminous parts of galaxies. Because there is a great amount of luminous stellar material in the cores of spiral galaxies and a deficit at large radii, the general consensus is that the dark matter halo component is not described by an isothermal sphere, but a model with a monotonically rising $\gamma(r)$ such that $\gamma(r) \ll 2$ inside the baryon dominated core, and $\gamma(r) \gg 2$ outside several Einstein radii.

Burkert [20] fitted a sample of several dark matter dominated dwarf galaxies employing a phenomenologically modified universal fitting formula,

$$\rho(r) = \frac{\rho_0 r_s^3}{(r+r_s)(r^2+r_s^2)}. \tag{8}$$

which, as in the case of the isothermal sphere of Equation (4), approximates a constant density core, $\gamma \to 0$ at $r \ll r_s$ – instead of a divergent $\gamma = 1$ core – but otherwise agrees with the NFW profile, with $\gamma \to 3$ at $r >> r_s$. Spherically integrating the Burkert model of Equation (8), one obtains an analytic fitting formula for the mass of dark matter,

$$M(r) = \pi\rho_0 r_s^3 \left\{ \ln(r^2+r_s^2) + 2\ln(r+r_s) - 4\ln(r_s) - 2\tan^{-1}(r/r_s) \right\}, \tag{9}$$

which diverges logarithmically, for $r \gg r_s$.

Observationally, when including baryons, cores of galaxies and clusters of galaxies are better fitted by a generalized fitting formula to account for more or less cuspy cores. Zhao [21] hypothesized that the NFW fitting formula must be broadened to account for the basic observed features of galactic dynamics, including less cuspy cores:

$$\rho(r) = \frac{\rho_0 r_s^m}{r^n \left(r^\alpha + r_s^\alpha\right)^{(m-n)/\alpha}}, \tag{10}$$

where (α, m, n) are free parameters. The NFW fitting formula of Equation (6) corresponds to Equation (10) with an inner cusp with logarithmic slope $n = 1$, an outer corona with logarithmic slope $m = 3$, and a "turnover" exponent of $\alpha = 1$. Syer and White [22] argued that the existence of a $\gamma \ll 1$ core is inconsistent with the hierarchical formation scenario of dark halos, which are much more likely to result in cuspy central density distributions. The least cuspy fitting formula, the isothermal spheres of Equations (4) and (5) correspond to Equation (10) with a constant density inner core with logarithmic slope $n = 0$, an outer corona with logarithmic slope $m = 2$, and a "turnover" exponent of $\alpha = 2$. Although Burkert's fitting formula of Equation (8) cannot be expressed in the form of Equation (10), it does bridge the constant density, $\gamma \to 0$, core behavior of the isothermal sphere with the $\gamma \to 3$ large r behavior of the NFW profile.

Mechanisms involving angular momentum transfer and dynamical friction may work to flatten the cusp in the cores of galaxies and clusters by dissipating the energy of the baryonic component, possibly overwhelming the reciprocal effect of adiabatic compression [23, 24, 25, 26, 27, 28]. Sand et al. [29, 30] used a spectroscopic survey of gravitational arcs in a sample of clusters with stellar velocity dispersion data, and found shallower dark matter profiles in the central regions. Newman et al. [31] used weak and strong lensing data, coupled with resolved stellar velocity dispersion data, to provide a

detailed analysis of the baryonic and dark matter distribution in the lensing cluster Abell 611 between 3 kpc to 3.25 Mpc, and found shallower dark matter profiles in the central regions. Zitrin and Broadhurst [32], Zitrin et al. [33] used strong galaxy lensing around the most massive X-ray luminous cluster systems, producing surface density maps with shallow mass profiles due to relatively unconcentrated dark matter halos, with steeper profiles in the center due to the contribution of central galaxies.

A simple choice for the dark matter halo, corresponding to Equation (10), which allows for the observed baryon distributions, without adding to the cusp problem, and fits both galaxy rotations curves and X-ray cluster mass profiles across all scales [14], is the *Brownstein profile*:

$$\rho(r) = \frac{\rho_0 r_s^3}{r^3 + r_s^3}, \tag{11}$$

with a constant density inner core with logarithmic slope $n = 0$, an outer corona with logarithmic slope $m = 3$, and a turnover exponent of $\alpha = 3$. Without the core of baryons, Equation (11) approximates a constant density core, $\gamma \to 0$ at $r \ll r_s$ – instead of a divergent $\gamma = 1$ core – but otherwise agrees with the NFW profile, with $\gamma \to 3$ at $r >> r_s$. Spherically integrating the density profile of Equation (11), one obtains the *Brownstein fitting formula* for the mass of dark matter,

$$M(r) = \frac{4}{3}\pi\rho_0 r_s^3 \ln\left(1 + \frac{r^3}{r_s^3}\right), \tag{12}$$

which diverges logarithmically, for $r \gg r_s$. The concentration parameter,

$$c \equiv r_{200}/r_s, \tag{13}$$

where r_{200} is the radius at which the dark matter density is 200 times the Λ-CDM critical density, follows from

$$\frac{M(x)}{M_{200}} = \frac{\ln(1 + c^3 x^3)}{\ln(1 + c^3)}, \tag{14}$$

where $x = r/r_{200}$, M_{200} is the mass of dark matter enclosed by a sphere of radius r_{200}, and $M(x)$ is the mass of dark matter enclosed by a sphere of radius x in units of r_{200}.

The surface mass density for the profile of Equation (11) is calculated by projecting the radial coordinate, $r = \sqrt{R^2 + z^2}$, and integrating the density along the line-of-sight, z:

$$\Sigma(R) = 2\rho_0 r_s^3 \int_0^\infty \frac{dz}{(R^2 + z^2)^{\frac{3}{2}} + r_s^3}, \tag{15}$$

where R is in the plane perpendicular to the line-of-sight, z. The surface mass density is a maximum at $R = 0$:

$$\Sigma(0) = \frac{4\rho_0 r_s \sqrt{3}\pi}{9}, \tag{16}$$

and then drops off as

$$\Sigma(R) = \frac{2\rho_0 r_s^3}{R^2} \quad R > r_s. \tag{17}$$

Utilizing the form of the power-law of Equation (3), the power-law index of the profile of Equation (11) is minus the logarithm slope

$$\gamma(r) = -\frac{d\ln\rho(r)}{d\ln r} = \frac{3r^3}{r^3 + r_s^3}. \tag{18}$$

The central density, $\rho(0) = \rho_0$, is finite and may be written in terms of the Λ-CDM critical density, $\rho_c(z)$, and the dark matter central overdensity, δ_0,

$$\rho_0 = \rho_c(z)\delta_0, \tag{19}$$

where z is the redshift. Moreover, the dark matter density at $r = r_s$ is one-half the central density,

$$\rho(r_s) = \frac{1}{2}\rho_0, \tag{20}$$

and the power-law index of Equation (18) is

$$\gamma(r_s) = 3/2, \tag{21}$$

which is the intermediate value between the inner core with logarithmic slope $\gamma \to 0$, and outer corona with logarithmic slope $\gamma \to 3$. This means that the halo's constant density core is limited to the region $r < r_s$, where baryons dominate the galaxy, which is important for N-body simulations.

Although the baryons represent a small fraction of the total, including their distribution is dynamically important in the cores of galaxies and rich clusters, and the combination of visible baryons and the dark matter mass of Equation (12) produces a mass function which correctly fits galaxy rotation curves, and X-ray cluster mass profiles, as shown in §3.1 and §3.2, which is neither the case for the pseudo-isothermal sphere nor the NFW profile. Using cosmological N-body simulations, Davé et al. [34] showed that generalizations of the NFW profile, such as Equation (11), may alleviate the problems arising from the singular NFW profile due to a shallower core, and Wyithe et al. [35] showed that generalized gravitational lenses may resolve further observational discrepancies.

3. THE GLOBAL MASS TO LIGHT RATIO

The missing mass problem in galaxies and clusters of galaxies is best quantified by calculating the spherically averaged global mass to light ratio,

$$\Upsilon(r) = M(r)/L(r), \tag{22}$$

where $M(r)$ is the dynamically inferred mass inside a sphere of radius, r, and $L(r)$ is the observed luminosity integrated over the same region. We assume that the baryon mass to light ratio,

$$\Upsilon_{\rm baryon} = M_{\rm baryon}(r)/L(r), \tag{23}$$

is constant for a particular object, and so we may write the global mass to light ratio,

$$\Upsilon(r) = \Upsilon_{\text{baryon}}\Gamma(r), \tag{24}$$

in terms of the dynamic mass ratio:

$$\Gamma(r) \equiv M(r)/M_{\text{baryon}}(r). \tag{25}$$

The dynamic mass is computed according to the Newtonian force law of Equation (1), in which case a halo of dark matter is required to explain the missing mass:

$$M(r) = M_{\text{baryon}}(r) + M_{\text{halo}}(r), \tag{26}$$

as demonstrated in §3.1.1, or the dynamic mass is computed according to a particular modified theory of gravity according to Equation (25),

$$M(r) = \Gamma(r)M_{\text{baryon}}(r), \tag{27}$$

where the dynamic mass factor,

$$\Gamma(r) = G(r)/G_N, \tag{28}$$

depends on the particular modified gravity acceleration law:

$$a(r) = -G(r)M_{\text{baryon}}(r)/r^2. \tag{29}$$

The form of Equation (29) can be applied to any modified theory of gravity which deviates from the Newtonian inverse square law, including Milgrom's MOND [36, 37] as shown in §3.1.2, and Moffat's MOG [38, 39, 40, 41, 42, 43] as shown in §3.1.3.

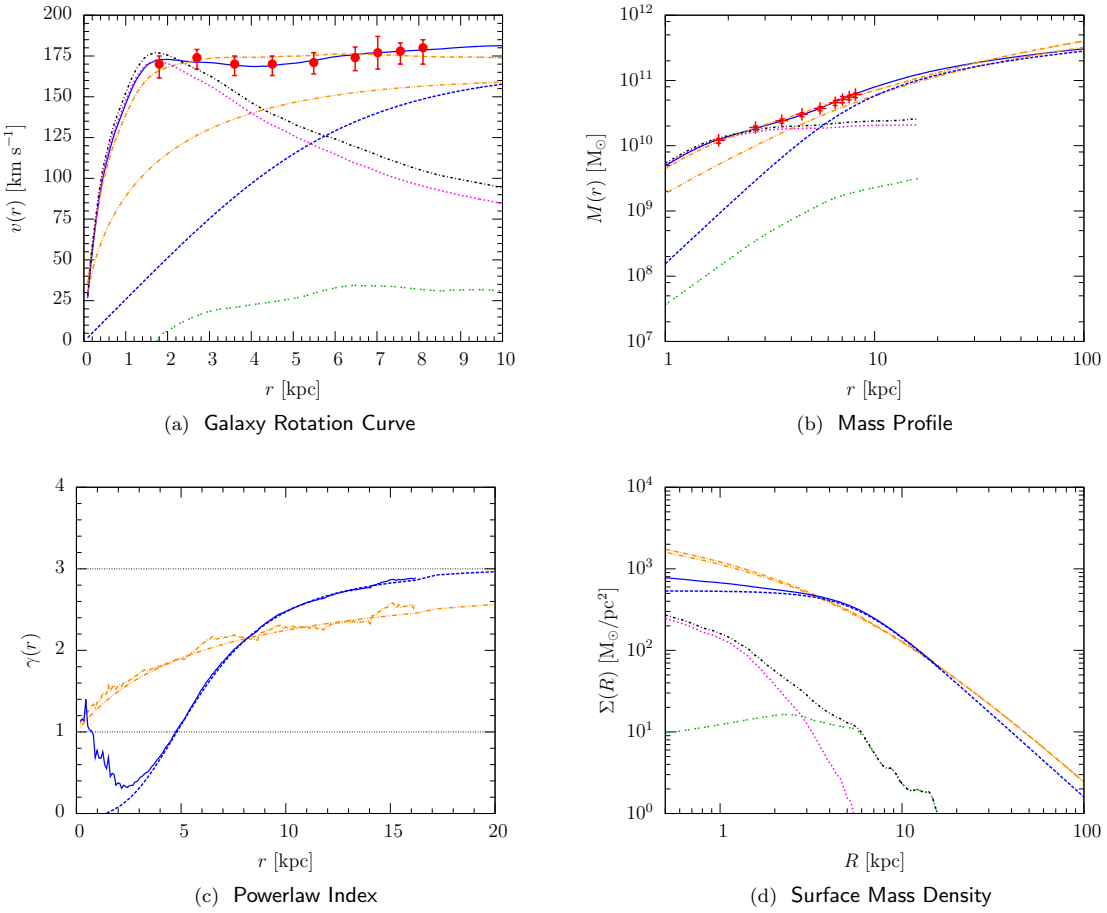
3.1. Galaxy Rotation Curves

3.1.1. Dark Matter

Every galaxy from the highest to lowest in surface brightness demonstrates a missing mass problem which increases with orbital distance from the center of the galaxy, and is greatest at the outermost extent of each galaxy, increasing monotonically with orbital distance. In Brownstein [14, Chapter 4], 29 HSB and LSB galaxy rotation curves were fitted with high-resolution Spitzer space telescope photometry using the dark matter fitting formula of Equation (11), producing excellent best fits to all of the galaxies including the *dwarf galaxies*, with consistent stellar mass to light ratios, showing a statistically significant reduction of the χ^2 per degree test of freedom in $\sim 90\%$ of the galaxies, over that of the NFW profile of Equation (6).

Figure 1a shows a comparison of galaxy rotation curves for the HSB galaxy UGC 6973, with the Brownstein model of Equation (11), shown in blue, and the NFW model of Equation (6) shown in yellow. All of the orphan features (kinks in the galaxy

v(r) data	red filled circles with error bars	HI gas	green dot-dotted line
M(r) data	red crosses with error bars	Stellar disk	magenta dotted line
Baryons (no Dark Matter)	black short-dash-dotted line		
Brownstein with Baryons	blue solid line	Brownstein halo	blue dashed line
NFW with Baryons	yellow dash-dotted line	NFW halo	yellow dash-dot-dotted line



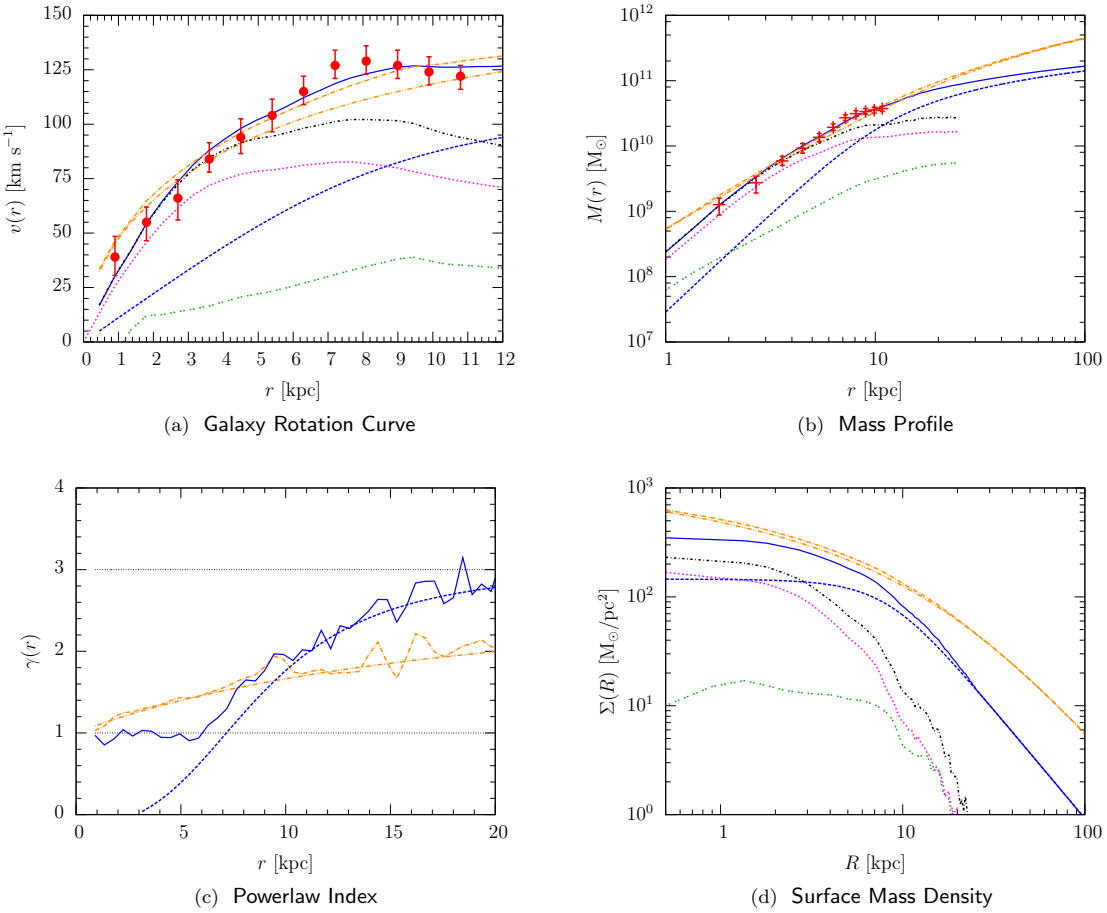
(a) Galaxy Rotation Curve (b) Mass Profile

(c) Powerlaw Index (d) Surface Mass Density

FIGURE 1. High surface brightness (HSB) dwarf spiral galaxy UGC 6973 [14, Chapter 4].

rotation curve) were correctly repatriated to the luminous components. The integrated mass profiles, $M(r)$, demonstrate the transition between the baryon dominated core and the dark matter dominated halo, as shown in Figure 1b. Both models models produce logarithmically divergent mass functions. However, the Brownstein model requires a much less massive halo and allows for a physically reasonable stellar mass to light ratio, consistent across the entire sample. Every galaxy differed strongly from a single power-law density profile, $\rho \propto r^{-\gamma}$. However, the NFW model requires a significantly smaller stellar mass to light ratio, so that the variation in the logarithmic density slope, $\gamma(r)$ according to Equation (3), shown in Figure 1c, is approximately the same with or without baryons. Conversely, including the HI (and He) gas and stellar disk photometry led to a *baryon corrected* power-law index, $\gamma(r)$, varying between 1 (at small r, within

v(r) data	red filled circles with error bars	HI gas	green dot-dotted line
M(r) data	red crosses with error bars	Stellar disk	magenta dotted line
Baryons (no Dark Matter)	black short-dash-dotted line		
Brownstein with Baryons	blue solid line	Brownstein halo	blue dashed line
NFW with Baryons	yellow dash-dotted line	NFW halo	yellow dash-dot-dotted line



(a) Galaxy Rotation Curve (b) Mass Profile

(c) Powerlaw Index (d) Surface Mass Density

FIGURE 2. Low surface brightness (LSB) dwarf spiral galaxy NGC 4010 [14, Chapter 4].

the baryon dominated core) and 3 (at large distances, within the dark matter halo). Surface mass density maps, $\Sigma(R)$, were produced for each galaxy as predictions for gravitational lensing studies, as shown for UGC 6973 in Figure 1d. Predictions for the Brownstein model differ from the NFW profile, which is everywhere dark matter dominated. However, like the NFW profile, the Brownstein model including baryons is nowhere flat.

Overall, the Brownstein model of Equation (11) provided very low χ^2 fits for each of the 19 HSB galaxies and each of the 10 LSB galaxies. This was not the case for the NFW model of Equation (6), which either fit the galaxy rotation curves very well, or failed to provide χ^2 best-fits for any positive definite stellar mass to light ratio. Figure 2a shows a comparison of galaxy rotation curves for the LSB dwarf galaxy NGC 4010, with the

Brownstein model of Equation (11), shown in blue, and the NFW model of Equation (6) shown in yellow. The Brownstein model repatriated all of the orphan features by requiring a stellar mass to light ratio consistent across the sample, whereas the NFW model required a nearly vanishing stellar mass to light ratio.

The integrated mass profiles, $M(r)$, demonstrate the transition between the baryon dominated core and the dark matter dominated halo, as shown in Figure 2b for the Brownstein model, whereas the the mass budget in the NFW model is everywhere dominated by dark matter. Remarkably, including the HI (and He) gas and stellar disk photometry led to a *baryon corrected* power-law index, $\gamma(r)$, for the Brownstein model, varying between 1 (at small r, within the baryon dominated core) and 3 (at large distances, within the dark matter halo), nearly identical to the NFW model which has completely suppressed the baryons by requiring a nearly vanishing stellar mass to light ratio, as shown for NGC 4010 in Figure 2c. As in the case of UGC 6973, the predicted surface density maps, $\Sigma(R)$, for gravitational lensing studies, as shown for NGC 4010 in Figure 2d show a baryon enhanced core for the Brownstein model, whereas the NFW model is everywhere dark matter dominated.

3.1.2. Milgrom's Modified Newtonian Dynamics (MOND)

The strong equivalence principle may be violated by preferred space-time fields. Milgrom [36] challenged the hidden mass hypothesis and introduced a nonrelativistic modification of Newtonian dynamics (MOND) at small accelerations, $a < a_0$, whereupon the gravitational acceleration of a test particle is given by

$$a\mu\left(\frac{a}{a_0}\right) = a_{\rm N}, \tag{30}$$

where $\mu(x)$ is a function that interpolates between the Newtonian regime, $\mu(x) = 1$, when $x \gg 1$ and the MOND regime, $\mu(x) = x$, when $x \ll 1$. Milgrom [37] introduced the interpolating function normally used for galaxy fitting,

$$\mu(x) = \frac{x}{\sqrt{1+x^2}}, \tag{31}$$

where

$$x \equiv x(r) = \left|\frac{\nabla\Phi(r)}{a_0}\right| = \left|\frac{a(r)}{a_0}\right|, \tag{32}$$

and determined that $a_0 \approx cH_0/6$, implying a cosmological connection to MOND:

$$a_0 = 1.0 \times 10^{-8}\ \text{cm s}^{-2}. \tag{33}$$

Substituting Equations (31) and (32) into Equation (30) gives,

$$\frac{a(r)^2}{\sqrt{a(r)^2 + a_0^2}} = a_{\rm N}(r), \tag{34}$$

which has the solution,

$$a(r) = a_0 \sqrt{\frac{1}{2}\left(\frac{a_N(r)}{a_0}\right)^2 + \sqrt{\frac{1}{4}\left(\frac{a_N(r)}{a_0}\right)^4 + \left(\frac{a_N(r)}{a_0}\right)^2}}, \tag{35}$$

written in terms of the Newtonian acceleration of a test particle at a separation, r,

$$a_N(r) = \frac{GM(r)}{r^2}, \tag{36}$$

where $M(r)$ is the baryonic mass integrated within a sphere of radius, r. Substituting Equation (36) into Equation (30), the MOND acceleration law can be written,

$$a(r) = \frac{1}{\mu(r)} \frac{G_N M(r)}{r^2}, \tag{37}$$

and therefore MOND can be interpreted as gravity theory with a varying gravitational coupling according to Equation (29), with

$$G(r) = \frac{G_N}{\mu(r)}, \tag{38}$$

and a dynamic mass factor, as defined by Equation (28), with

$$\Gamma(r) = \frac{1}{\mu(r)}, \tag{39}$$

where $\Gamma(r) \sim 1$ in the Newtonian regime and $\Gamma(r) \gg 1$ in the MOND regime. It is important to note that MOND has a classical instability in the deep MOND regime corresponding to $\mu \to 0$ which leads to a divergent gravitational coupling, $\Gamma(r) \to \infty$, and that MOND violates the strong equivalence principle for all $\mu \neq 1$.

For gravity fields interior to galaxies and clusters of galaxies, the accelerations are sufficiently small that the MOND interpolating function, $\mu(x) \gg 1$, so that the Newtonian dynamic mass determined by MOND is much larger than the actual mass visible in the system, so that MOND may be an alternative to dark matter in these systems.

3.1.3. Moffat's Modified Gravity (MOG)

Clayton [44] showed that Moffat's massive nonsymmetric gravity theory (NGT) becomes identical to a Kalb-Ramond-Proca field [45] with an additional curvature coupling term when considered as a perturbation about a Ricci-flat background. Moffat [46, 47], Moffat and Sokolov [48] determined that in the weak-field approximation of the massive NGT relevant to galaxy dynamics, a range dependent Yukawa-type, fifth force [49] emerges in addition to the Newtonian $1/r^2$ central force; and asserted that this additional potential due to the interaction of the field structure with matter in the halos of

galaxies can explain galaxy rotation curves. van Nieuwenhuizen [50] found that the only massive antisymmetric tensor or vector fields free of ghosts, tachyons and higher-order poles in the propagator for linearized gravitation are the Maxwell-Proca fields. Isenberg and Nester [51] determined that only Maxwell fields, Proca fields, and purely longitudinal vector fields are free of instability when minimally coupled to gravity. Damour et al. [52] analyzed a class of physically consistent and ghost-free nonsymmetric gravity models with finite range massive gauge bosons, such as Moffat's massive NGT, and showed that the Kalb-Ramond-Proca field leads to minuscule – as yet unmeasured – deviations from Newtonian gravity at terrestrial scales consistent with stringent experimental bounds on possible violations, and the field may acquire *gravitational* strength at sufficiently large astrophysical scales.

Moffat [38] introduced Metric skew-tensor gravity (MSTG), where the modified acceleration law results from coupling the Kalb-Ramond-Proca field of the massive NGT to Einstein's metric. At astrophysical scales, the emergent low energy Yukawa meson is the only feature of the massive NGT left in MSTG to explain galaxy rotation curves.

Moffat [39] introduced the scalar-tensor-vector gravity (STVG) theory by adding a simpler, massive and self-coupled Maxwell-Proca field coupled to matter and gravity. The STVG effectively captures the fifth force due to a weak-field emergent Yukawa meson, simulating the predictions of massive NGT and MSTG, to leading order.

Moffat's modified gravity (MOG) theories have an acceleration law with a varying gravitational coupling according to Equation (29), with

$$G(r) = G_N\left\{1+\alpha\left[1-\exp(-\mu r)(1+\mu r)\right]\right\}, \tag{40}$$

and a dynamic mass factor, as defined by Equation (28), with

$$\Gamma(r) = \left\{1+\alpha\left[1-\exp(-\mu r)(1+\mu r)\right]\right\}, \tag{41}$$

In the absence of analytic solutions, MSTG requires phenomenological input from the Tully-Fisher relation, leading to the parameters, M_0 and r_0,

$$\alpha = \sqrt{\frac{M_0}{M}}, \tag{42}$$

$$\mu = 1/r_0, \tag{43}$$

used to fit galaxy rotation curves and X-ray cluster masses [40, 41, 42].

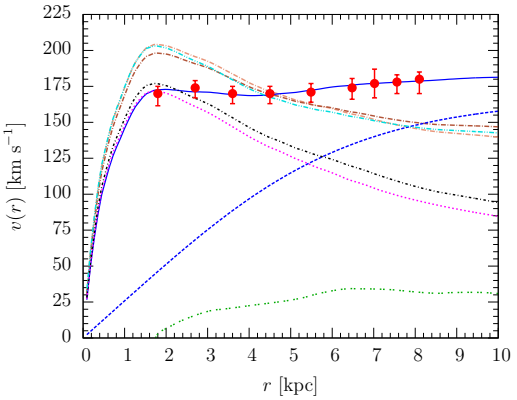
Moffat and Toth [43] showed that phenomenological input can be avoided in STVG by integrating the equations of motion in the weak-field, spherically symmetric limit, leading to solutions for α and μ as functions of the mass M,

$$\alpha = \frac{M}{\left(\sqrt{M}+E\right)^2}\left(\frac{G_\infty}{G_N}-1\right), \tag{44}$$

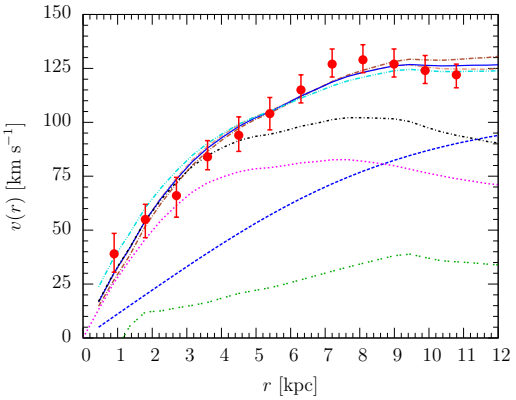
$$\mu = \frac{D}{\sqrt{M}}, \tag{45}$$

where the parameters, D, E and G_∞, are constants of integration to be determined.

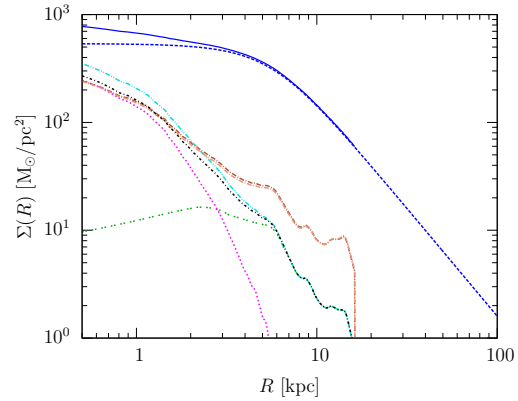
v(r) data	red filled circles with error bars	HI gas	green dot-dotted line
Baryons (no Dark Matter)	black short-dash-dotted line	Stellar disk	magenta dotted line
Brownstein with Baryons	blue solid line	Brownstein halo	blue dashed line
Moffat's STVG	dark brown dash-dotted line	Moffat's MSTG	light brown dash-dotted line
Milgrom's MOND	cyan dash-dot-dotted line		



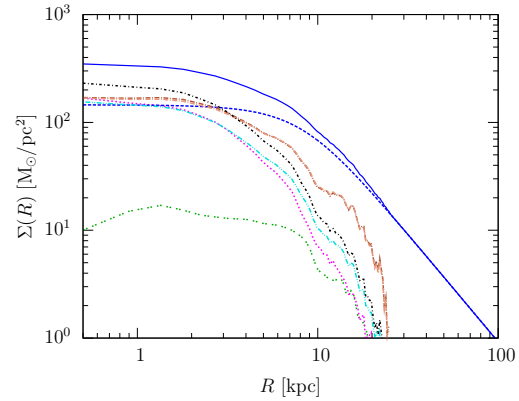
(a) UGC 6973 Galaxy Rotation Curve



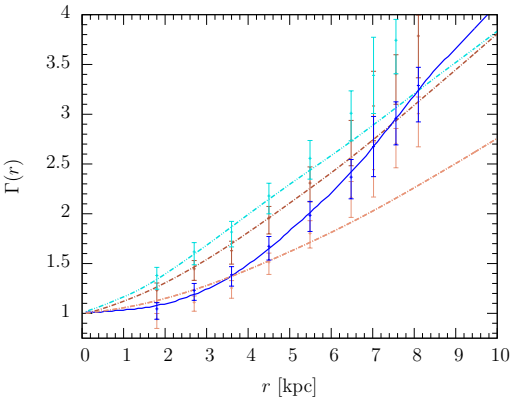
(b) NGC 4010 Galaxy Rotation Curve



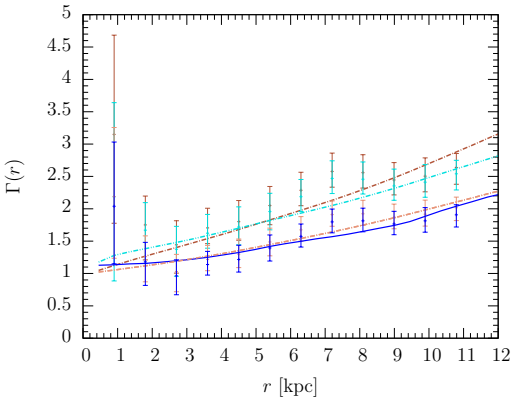
(c) UGC 6973 Surface Mass Density



(d) NGC 4010 Surface Mass Density



(e) UGC 6973 Dynamic Mass Factor



(f) NGC 4010 Dynamic Mass Factor

FIGURE 3. Comparison of The Brownstein dark matter model to Milgrom's MOND, and to Moffat's MSTG and STVG theories for the high surface brightness (HSB) dwarf spiral galaxy UGC 6973 [left panels], and low surface brightness (LSB) dwarf spiral galaxy NGC 4010 [right panels] [14, Chapter 4].

Brownstein [14, Chapter 4] fitted a sample of 19 HSB and 10 LSB galaxies, comparing the Brownstein dark matter model to Milgrom's MOND, and Moffat's MOG.

To compute the MOND universal acceleration parameter, the MOND acceleration law of Equation (35) was applied with a variable a_0 yielding the subsample averages:

$$a_0 = \begin{cases} (1.34 \pm 0.66) \times 10^{-8} \text{ cm s}^{-2} & \texttt{HSB galaxy subsample}, \\ (1.02 \pm 0.78) \times 10^{-8} \text{ cm s}^{-2} & \texttt{LSB galaxy subsample}. \end{cases} \tag{46}$$

Because of the gross uncertainty in the averaged results, all galaxy rotation curves were refitted using Equation (33) universally.

The computation of the universal parameters in Moffat's MOG was performed by allowing M_0 and r_0 in MSTG, and D, E, and G_∞ in STVG to vary across the sample of 29 HSB and LSB galaxies, with the overall best-fitting results for each theory, respectively:

$$D = (6.4 \pm 0.2)\sqrt{M_\odot}\text{pc}^{-1}, \quad E = (28.4 \pm 7.9) \times 10^3 M_\odot^{1/2}, \quad G_\infty = (24 \pm 18) G_N, \tag{47}$$

$$M_0 = (98.6 \pm 21.6) \times 10^{10} M_\odot, \quad r_0 = (16.4 \pm 6.1)\,\text{kpc}. \tag{48}$$

The galaxy rotation curves were subsequently refitted as **one parameter** best-fits by varying only the stellar mass-to-light ratio, Υ, using Equations (47) and (48) universally.

Comparison of the best-fit galaxy rotation curves for UGC 6973, plotted in Figure 3a, shows that Moffat's MOG theories with universal parameters across the subsample perform as well as Milgrom's MOND theory, but not as well as the Brownstein dark matter model with variable dark matter halo parameters, ρ_0 and r_s. However, comparison of the best-fit galaxy rotation curves for NGC 4010, plotted in Figure 3b, shows that Milgrom's MOND and Moffat's MOG with universal parameters do as well as the Brownstein dark matter model, whereas the NFW model plotted in Figure 2b does not provide a comparable low χ^2 fit. Perhaps most significantly, the total mass with dark matter in the Brownstein model showed the least scatter of any of the Tully-Fisher relations, even when compared to Milgrom's MOND and Moffat's MSTG, which receive phenomenological input from the Tully-Fisher relation.

In order to better distinguish the models, it is necessary to perform precise strong gravitational lensing measurements of the total surface mass density, $\Sigma(r)$, plotted in Figure 3c and Figure 3d for UGC 6973 and NGC 4010, respectively.

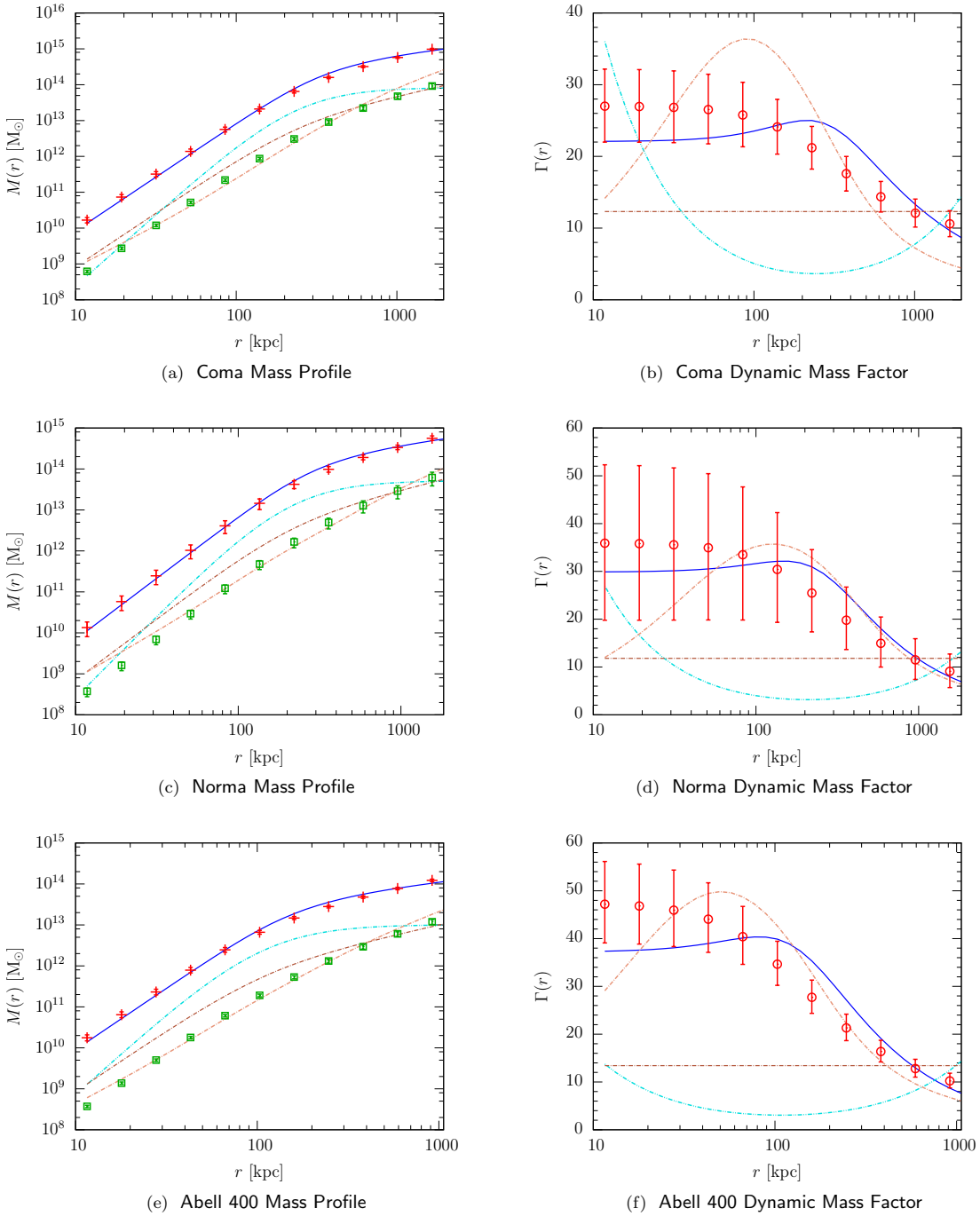
Nevertheless, it is remarkable that every model provides a similar picture in terms of the dynamic mass factor of Equation (25), plotted in Figure 3d and Figure 3e for UGC 6973 and NGC 4010, respectively. For the complete sample of 29 galaxies, the dynamic mass factor varies as,

$$\Gamma(r) \to 1 \text{ as } r \to 0, \tag{49}$$

$$3 \lesssim \Gamma_{\max}(r) \lesssim 7 \text{ as } r \to r_{\max}. \tag{50}$$

Therefore, we may conclude that every galaxy has a central core of baryons, outside of which there must be either a halo of dark matter, or a modified gravity region, which dominates the gravitational potential. This trend common among galaxy rotation curves is not seen in X-ray cluster masses, as detailed in §3.2, where the dynamic mass factor varies according to Equation (51), in contrast to Equations (49) and (50).

Newtonian dynamic mass	red crosses	Dynamic mass factor	red open circles
ICM gas mass	green open squares		
Brownstein with Baryons	blue solid line	Milgrom's MOND	cyan dash-dot-dotted line
Moffat's STVG	dark brown dash-dotted line	Moffat's MSTG	light brown dash-dotted line



(a) Coma Mass Profile

(b) Coma Dynamic Mass Factor

(c) Norma Mass Profile

(d) Norma Dynamic Mass Factor

(e) Abell 400 Mass Profile

(f) Abell 400 Dynamic Mass Factor

FIGURE 4. Comparison of The Brownstein dark matter model to Milgrom's MOND, and to Moffat's MSTG and STVG theories for the Coma, Norma and Abell 400 X-ray clusters, showing mass profiles [left panels], and dynamic mass factors [right panels] [14, Chapter 5].

3.2. X-ray Cluster Imaging

Every X-ray emitting cluster of galaxies from 10 million degrees to 170 million degrees demonstrates a missing mass problem which is greatest at the center, and decreases with orbital distance from the center of the cluster, as seen by the variation in the dynamic mass factor of Equation (25), which varies as,

$$7 \lesssim \Gamma_{\max}(r) \lesssim 500 \text{ as } r \to 0, \tag{51}$$

and $\Gamma(r)$ decreases monotonically with orbital distance. Therefore, every cluster of galaxies is everywhere dominated by either dark matter, or modified gravity everywhere dominates the gravitational potential, particularly towards the center. Moreover, the missing mass problem is considerably stronger in clusters of galaxies, than galaxies.

Brownstein [14, Chapter 5] fitted a sample of 10 X-ray clusters of galaxies taken from a sample of 106 X-ray clusters [53, 54] considered in [41], comparing the Brownstein dark matter model of Equation (11) to Milgrom's MOND, and Moffat's MOG theories. The Brownstein dark matter model provided excellent fits with low χ^2 to every X-ray mass profile in the sample, as shown in Figure 4 for Coma, Norma and Abell 400, for example, whereas attempts to fit cluster mass distributions to NFW profiles led to large uncertainties due to a parameter degeneracy between the central density parameter, ρ_0, and the scale radius, r_s, which prevented the computation of the best-fit ρ_0 and r_s from converging, regardless of the χ^2. Without numerical convergence, the NFW results either over-predict the density at the core or under-predict the total mass.

Although MOND predicts convergent mass profiles, none matched any of the observed distributions to low χ^2. Treating the MOND universal acceleration, a_0, as a variable parameter did little to improve the χ^2 because the dynamic mass factor predicted by MOND at the scale of clusters of galaxies does not correctly track the data, as shown in the rightmost panels of Figure 4, whereas the Brownstein dark matter model follows the data across the entire cluster within one standard deviation, for every cluster of galaxies, tracking the data with a large, and nearly constant, $\Gamma(r)$, which decreases monotonically at large radii.

Moffat's MSTG was not able to provide low χ^2 fits to any of the clusters of galaxies with the universal parameters of Equation (48), determined from fitting the galaxy rotation curves. In order to better determine the scale dependence of the parameters, it is reasonable to treat the MSTG mass and range parameters as variable and to perform two-parameter best-fits to the X-ray gas masses of the sample of clusters of galaxies. MSTG was able to provide moderate to low χ^2 fits to all of the X-ray mass profiles by allowing variable M_0 and r_0, as shown in the leftmost panels of Figure 4. The dynamic mass factors in MSTG, plotted in the rightmost panels of Figure 4, are similar to the Brownstein model at large distances from the cluster center, but do not track the data with the observed constant Γ needed to fit the core. Whereas tabulation of the best-fit MSTG parameters provided no statistical support for universal constants, there was strong statistical support that the MSTG parameters are scale-dependent.

Moffat's STVG was not able to provide low χ^2 fits to any of the clusters of galaxies with the universal parameters of Equation (47), determined from fitting the galaxy rotation curves. Whereas every weak fitting STVG one-parameter best-fit by a variable

stellar mass-to-light ratio, Υ, show dramatic improvement and reduction in the reduced χ^2/ν statistic using a four-parameter best-fit including variable parameters, the tabulation of D, E, and G_∞ provided no statistical support that the STVG integration constants are universal. However, for values of D sufficiently large and values of E sufficiently small, the STVG gravitational coupling and dynamic mass factor of Equations (40) and (41), respectively, simplify to their asymptotic form,

$$G(r) = G_\infty, \quad \Gamma(r) = G_\infty/G_N, \tag{52}$$

and are independent of r. Therefore, in order to determine the scale dependence of the STVG asymptotic coupling, it is reasonable to treat G_∞ as variable and to perform one-parameter best-fits to the X-ray gas masses of the sample of clusters of galaxies, shown in the leftmost panels of Figure 4. Because the STVG dynamic mass factor is constant, whereas the data is only constant in the core and then monotonically decreasing thereafter, the spherically symmetric STVG model does not correctly track the missing mass problem in clusters of galaxies, as shown in the rightmost panels of Figure 4.

4. CONCLUSIONS

The Brownstein model of Equation (11) does provide a universal fitting formula from which excellent fits, with low χ^2, are obtained across the scales from dwarf galaxies to the largest cluster of galaxies, provided the observed baryons are not neglected in the mass budget. This is an important step in imaging dark matter halos, particularly because the NFW model of Equation (6) fails at the low mass scale of dwarf galaxies, and at the high mass scale of X-ray clusters of galaxies. Conversely, none of the alternatives to dark matter considered in [14] fit the missing mass problem with common universal parameters across the scales from dwarf galaxies to clusters of galaxies. This suggests that if the equivalence principle is violated at astrophysical scales with sufficient strength to explain the missing matter problem, then the process must be dynamical, allowing for running coupling constants, explaining the success of Moffat's MOG to fit the data at all scales with variable parameters.

REFERENCES

1. J. Aguirre, A. Amblard, A. Ashoorioon, C. Baccigalupi, A. Balbi, J. Bartlett, N. Bartolo, D. Benford, M. Birkinshaw, J. Bock, D. Bond, J. Borrill, F. Bouchet, M. Bridges, E. Bunn, E. Calabrese, C. Cantalupo, A. Caramete, C. Carbone, S. Chatterjee, S. Church, D. Chuss, C. Contaldi, A. Cooray, S. Das, F. De Bernardis, P. De Bernardis, G. De Zotti, J. Delabrouille, F. -Xavier Dsert, M. Devlin, C. Dickinson, S. Dicker, M. Dobbs, S. Dodelson, O. Dore, J. Dotson, J. Dunkley, M. C. Falvella, D. Fixsen, P. Fosalba, J. Fowler, E. Gates, W. Gear, S. Golwala, K. Gorski, A. Gruppuso, J. Gundersen, M. Halpern, S. Hanany, M. Hazumi, C. Hernandez-Monteagudo, M. Hertzberg, G. Hinshaw, C. Hirata, E. Hivon, W. Holmes, W. Holzapfel, W. Hu, J. Hubmayr, K. Huffenberger, K. Irwin, M. Jackson, A. Jaffe, B. Johnson, W. Jones, M. Kaplinghat, B. Keating, R. Keskitalo, J. Khoury, W. Kinney, T. Kisner, L. Knox, A. Kogut, E. Komatsu, A. Kosowsky, J. Kovac, L. Krauss, H. Kurki-Suonio, S. Landau, C. Lawrence, S. Leach, A. Lee, E. Leitch, R. Leonardi, J. Lesgourgues, A. Liddle, E. Lim, M. Limon, M. Loverde, P. Lubin, A. Magalhaes, D. Maino, T. Marriage, V. Martin, S. Matarrese, J. Mather, H. Mathur, T. Matsumura, P. Meerburg, A. Melchiorri, S. Meyer, A. Miller,

M. Milligan, K. Moodley, M. Neimack, H. Nguyen, I. O'Dwyer, A. Orlando, L. Pagano, L. Page, B. Partridge, T. Pearson, H. Peiris, F. Piacentini, L. Piccirillo, E. Pierpaoli, D. Pietrobon, G. Pisano, L. Pogosian, D. Pogosyan, N. Ponthieu, L. Popa, C. Pryke, C. Raeth, S. Ray, C. Reichardt, S. Ricciardi, P. Richards, G. Rocha, L. Rudnick, J. Ruhl, B. Rusholme, C. Scoccola, D. Scott, C. Sealfon, N. Sehgal, M. Seiffert, L. Senatore, P. Serra, S. Shandera, M. Shimon, P. Shirron, J. Sievers, K. Sigurdson, J. Silk, R. Silverberg, E. Silverstein, S. Staggs, A. Stebbins, F. Stivoli, R. Stompor, N. Sugiyama, D. Swetz, A. Tartari, M. Tegmark, P. Timbie, M. Tristram, G. Tucker, J. Urrestilla, J. Vaillancourt, M. Veneziani, L. Verde, J. Vieira, S. Watson, B. Wandelt, G. Wilson, E. Wollack, M. Wyman, A. Yadav, G.-H. Yannick, O. Zahn, M. Zaldarriaga, M. Zemcov, and J. Zwart (2009), arXiv:0903.0902.

2. M. R. Nolta, J. Dunkley, R. S. Hill, G. Hinshaw, E. Komatsu, D. Larson, L. Page, D. N. Spergel, C. L. Bennett, B. Gold, N. Jarosik, N. Odegard, J. L. Weiland, E. Wollack, M. Halpern, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, and E. L. Wright, *Astrophys. J. Suppl. Series* **180**, 296–305 (2009), arXiv:0803.0593.
3. W. J. Percival, B. A. Reid, D. J. Eisenstein, N. A. Bahcall, T. Budavari, J. A. Frieman, M. Fukugita, J. E. Gunn, Ž. Ivezić, G. R. Knapp, R. G. Kron, J. Loveday, R. H. Lupton, T. A. McKay, A. Meiksin, R. C. Nichol, A. C. Pope, D. J. Schlegel, D. P. Schneider, D. N. Spergel, C. Stoughton, M. A. Strauss, A. S. Szalay, M. Tegmark, M. S. Vogeley, D. H. Weinberg, D. G. York, and I. Zehavi, *Mon. Not. Roy. Astron. Soc.* **401**, 2148–2168 (2010), arXiv:0907.1660.
4. N. Padmanabhan, and M. White, *Phys. Rev. D* **80**, 063508 (2009), arXiv:0906.1198.
5. E. A. Kazin, M. R. Blanton, R. Scoccimarro, C. K. McBride, and A. A. Berlind, *Astrophys. J.* **719**, 1032–1044 (2010), arXiv:1004.2244.
6. H. Lampeitl, R. C. Nichol, H. Seo, T. Giannantonio, C. Shapiro, B. Bassett, W. J. Percival, T. M. Davis, B. Dilday, J. Frieman, P. Garnavich, M. Sako, M. Smith, J. Sollerman, A. C. Becker, D. Cinabro, A. V. Filippenko, R. J. Foley, C. J. Hogan, J. A. Holtzman, S. W. Jha, K. Konishi, J. Marriner, M. W. Richmond, A. G. Riess, D. P. Schneider, M. Stritzinger, K. J. van der Heyden, J. T. Vanderplas, J. C. Wheeler, and C. Zheng, *Mon. Not. Roy. Astron. Soc.* **401**, 2331–2342 (2010), arXiv:0910.2193.
7. E. Komatsu, J. Dunkley, M. R. Nolta, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. Limon, L. Page, D. N. Spergel, M. Halpern, R. S. Hill, A. Kogut, S. S. Meyer, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright, *Astrophys. J. Suppl. Series* **180**, 330–376 (2009), arXiv:0803.0547.
8. M. Arnaud, H. Bohringer, C. Jones, B. McNamara, T. Ohashi, D. Patnaude, K. Arnaud, M. Bautz, A. Blanchard, J. Bregman, G. Chartas, J. Croston, L. David, M. Donahue, A. Fabian, A. Finoguenov, A. Furuzawa, S. Gallagher, Y. Haba, A. Hornschemeier, S. Heinz, J. Kaastra, W. Kapferer, G. Lamer, A. Mahdavi, K. Makishima, K. Matsushita, K. Nakazawa, P. Nulsen, P. Ogle, E. Perlman, T. Ponman, D. Proga, G. Pratt, S. Randall, G. Richards, K. Romer, M. Ruszkowski, R. Schmidt, R. Smith, H. Tananbaum, J. Vrtilek, and D. Worrall (2009), arXiv:0902.4890.
9. J. R. Primack, *Lectures at XIII Special Courses at Observatorio Nacional, Rio de Janeiro, Brazil.* (2009), arXiv:0909.2021.
10. J. F. Navarro, A. Ludlow, V. Springel, J. Wang, M. Vogelsberger, S. D. M. White, A. Jenkins, C. S. Frenk, and A. Helmi, *Mon. Not. Roy. Astron. Soc.* **402**, 21–34 (2010), arXiv:0810.1522.
11. J. Stadel, D. Potter, B. Moore, J. Diemand, P. Madau, M. Zemp, M. Kuhlen, and V. Quilis, *Mon. Not. Roy. Astron. Soc.* **398**, L21–L25 (2009), arXiv:0808.2981.
12. A. Burkert, *Astrophys. J. Lett.* **447**, L25 (1995), arXiv:astro-ph/9504041.
13. E. Zackrisson, N. Bergvall, T. Marquart, and G. Östlin, *Astron. & Astrophys.* **452**, 857–868 (2006), arXiv:astro-ph/0603523.
14. J. R. Brownstein, *Modified Gravity and the Phantom of Dark Matter*, Ph.D. thesis, University of Waterloo at Perimeter Institute for Theoretical Physics (2009), arXiv:0908.0040.
15. L. M. Voigt, and A. C. Fabian, *Mon. Not. Roy. Astron. Soc.* **368**, 518–533 (2006), arXiv:astro-ph/0602373.
16. J. P. Ostriker, P. J. E. Peebles, and A. Yahil, *Astrophys. J. Lett.* **193**, L1–L4 (1974).
17. K. G. Begeman, A. H. Broeils, and R. H. Sanders, *Mon. Not. Roy. Astron. Soc.* **249**, 523–537 (1991).
18. J. F. Navarro, C. S. Frenk, and S. D. M. White, *Astrophys. J.* **462**, 563–575 (1996), arXiv:astro-ph/9508025.

19. J. F. Navarro, C. S. Frenk, and S. D. M. White, *Astrophys. J.* **490**, 493–508 (1997), arXiv:astro-ph/9611107.
20. A. Burkert, *Astrophys. J. Lett.* **447**, L25–L28 (1995), arXiv:astro-ph/9504041.
21. H. Zhao, *Mon. Not. Roy. Astron. Soc.* **278**, 488–496 (1996), arXiv:astro-ph/9509122.
22. D. Syer, and S. D. M. White, *Mon. Not. Roy. Astron. Soc.* **293**, 337–342 (1998).
23. C. Nipoti, T. Treu, L. Ciotti, and M. Stiavelli, *Mon. Not. Roy. Astron. Soc.* **355**, 1119–1124 (2004), arXiv:astro-ph/0404127.
24. A. El-Zant, I. Shlosman, and Y. Hoffman, *Astrophys. J.* **560**, 636–643 (2001), arXiv:astro-ph/0103386.
25. A. A. El-Zant, Y. Hoffman, J. Primack, F. Combes, and I. Shlosman, *Astrophys. J. Lett.* **607**, L75–L78 (2004), arXiv:astro-ph/0309412.
26. E. Romano-Díaz, I. Shlosman, Y. Hoffman, and C. Heller, *Astrophys. J. Lett.* **685**, L105–L108 (2008), arXiv:0808.0195.
27. J. R. Jardel, and J. A. Sellwood, *Astrophys. J.* **691**, 1300–1306 (2009), arXiv:0808.3449.
28. A. Del Popolo, and P. Kroupa, *Astron. & Astrophys.* **502**, 733–747 (2009), arXiv:0906.1146.
29. D. J. Sand, T. Treu, G. P. Smith, and R. S. Ellis (2003), arXiv:astro-ph/0309465.
30. D. J. Sand, T. Treu, G. P. Smith, and R. S. Ellis, *Astrophys. J.* **604**, 88–107 (2004), arXiv:astro-ph/0310703.
31. A. B. Newman, T. Treu, R. S. Ellis, D. J. Sand, J. Richard, P. J. Marshall, P. Capak, and S. Miyazaki, *Astrophys. J.* **706**, 1078–1094 (2009), arXiv:0909.3527.
32. A. Zitrin, and T. Broadhurst, *Astrophys. J. Lett.* **703**, L132–L136 (2009), arXiv:0906.5079.
33. A. Zitrin, T. Broadhurst, Y. Rephaeli, and S. Sadeh, *Astrophys. J. Lett.* **707**, L102–L106 (2009), arXiv:0907.4232.
34. R. Davé, D. N. Spergel, P. J. Steinhardt, and B. D. Wandelt, *Astrophys. J.* **547**, 574–589 (2001), arXiv:astro-ph/0006218.
35. J. S. B. Wyithe, E. L. Turner, and D. N. Spergel, *Astrophys. J.* **555**, 504–523 (2001), arXiv:astro-ph/0007354.
36. M. Milgrom, *Astrophys. J.* **270**, 365–370 (1983a).
37. M. Milgrom, *Astrophys. J.* **270**, 371–383 (1983b).
38. J. W. Moffat, *JCAP05 003* (2005), arXiv:astro-ph/0412195.
39. J. W. Moffat, *JCAP03 004* (2006), arXiv:gr-qc/0506021.
40. J. R. Brownstein, and J. W. Moffat, *Astrophys. J.* **636**, 721–741 (2006), arXiv:astro-ph/0506370.
41. J. R. Brownstein, and J. W. Moffat, *Mon. Not. Roy. Astron. Soc.* **367**, 527–540 (2006), arXiv:astro-ph/0507222.
42. J. R. Brownstein, and J. W. Moffat, *Mon. Not. Roy. Astron. Soc.* **382**, 29–47 (2007), arXiv:astro-ph/0702146.
43. J. W. Moffat, and V. T. Toth, *Classical and Quantum Gravity* **26**, 085002 (2009), arXiv:0712.1796.
44. M. A. Clayton, *Journal of Mathematical Physics* **37**, 395–420 (1996), arXiv:gr-qc/9505005.
45. M. Kalb, and P. Ramond, *Phys. Rev. D* **9**, 2273–2284 (1974).
46. J. W. Moffat, "Nonsymmetric Gravitational Theory as a String Theory" (1995), arXiv:hep-th/9512018.
47. J. W. Moffat, *Journal of Mathematical Physics* **36**, 3722–3732 (1995).
48. J. W. Moffat, and I. Y. Sokolov, *Physics Letters B* **378**, 59–67 (1996), arXiv:astro-ph/9509143.
49. H. Yukawa, *Proc. Phys. Math. Soc. Jap.* **17**, 48–57 (1935).
50. P. van Nieuwenhuizen, *Nuclear Physics B* **60**, 478–492 (1973).
51. J. A. Isenberg, and J. M. Nester, *Annals of Physics* **107**, 56–81 (1977).
52. T. Damour, S. Deser, and J. McCarthy, *Phys. Rev. D* **47**, 1541–1556 (1993).
53. T. H. Reiprich, *Cosmological Implications and Physical Properties of an X-Ray Flux-Limited Sample of Galaxy Clusters*, Ph.D. thesis, Ludwig-Maximilians-Univ. (2001).
54. T. H. Reiprich, and H. Böhringer, *Astrophys. J.* **567**, 716–740 (2002), arXiv:astro-ph/0111285.

Making the Case for Conformal Gravity

Philip D. Mannheim

Department of Physics, University of Connecticut, Storrs, CT 06269, USA
philip.mannheim@uconn.edu

Abstract. We review some recent developments in the conformal gravity theory that has been advanced as a candidate alternative to standard Einstein gravity. As a quantum theory the conformal theory is both renormalizable and unitary, with unitarity being obtained because the theory is a *PT* symmetric rather than a Hermitian theory. We show that in the theory there can be no a priori classical curvature, with all curvature having to result from quantization. In the conformal theory gravity requires no independent quantization of its own, with it being quantized solely by virtue of its being coupled to a quantized matter source. Moreover, because it is this very coupling that fixes the strength of the gravitational field commutators, the gravity sector zero-point energy density and pressure fluctuations are then able to identically cancel the zero-point fluctuations associated with the matter sector. In addition, we show that when the conformal symmetry is spontaneously broken, the zero-point structure automatically readjusts so as to identically cancel the cosmological constant term that dynamical mass generation induces. We show that the macroscopic classical theory that results from the quantum conformal theory incorporates global physics effects that provide for a detailed accounting of a comprehensive set of 110 galactic rotation curves with no adjustable parameters other than the galactic mass to light ratios, and with the need for no dark matter whatsoever. With these global effects eliminating the need for dark matter, we see that invoking dark matter in galaxies could potentially be nothing more than an attempt to describe global physics effects in purely local galactic terms. Finally, we review some recent work by 't Hooft in which a connection between conformal gravity and Einstein gravity has been found.

Keywords: conformal gravity, quantum gravity, cosmological constant problem
PACS: 04.60.-m, 04.50.Kd, 04.90.+e

1. EINSTEIN GRAVITY: WHAT MUST BE KEPT

Following his development of special relativity, Einstein was faced with two immediate problems. The first was to make Newtonian gravity compatible with relativity, and the second was to develop a formalism in which not only uniformly moving observers but also accelerating ones would all be able agree on the same physics. While these two issues are logically independent (even in the absence of gravity one has to able to write Newton's second law of motion in an observer-independent way), by imposing general coordinate invariance and by identifying the spacetime metric $g_{\mu\nu}(x)$ as the gravitational field, Einstein was able to provide a solution to both problems simultaneously. In this formalism a central role is played by the Christoffel symbols

$$\Gamma^{\mu}_{\nu\sigma} = \frac{1}{2} g^{\mu\lambda} \left[\partial_{\nu} g_{\lambda\sigma} + \partial_{\sigma} g_{\lambda\nu} - \partial_{\lambda} g_{\nu\sigma}\right], \tag{1}$$

since in terms of them one can show that the path that minimizes the distance

$$ds^2 = g_{\mu\nu}(x) dx^{\mu} dx^{\nu} \tag{2}$$

between two points in any geometry (curved or flat) is the one that obeys the geodesic equation

$$\frac{d^2x^\mu}{ds^2}+\Gamma^\mu_{\nu\sigma}\frac{dx^\nu}{ds}\frac{dx^\sigma}{ds}=0. \tag{3}$$

The key feature of this equation is that even though neither of the two terms that appears in (3) is itself a true coordinate vector, the linear combination of them is, with the vanishing of their sum in any given coordinate system ensuring its vanishing in any other.

Even though the Christoffel symbols are not themselves true coordinate tensors, from them one can construct a quantity that is, viz. the Riemann tensor as defined by

$$R^\lambda{}_{\mu\nu\kappa}=\frac{\partial\Gamma^\lambda_{\mu\nu}}{\partial x^\kappa}+\Gamma^\lambda_{\kappa\eta}\Gamma^\eta_{\mu\nu}-\frac{\partial\Gamma^\lambda_{\mu\kappa}}{\partial x^\nu}-\Gamma^\lambda_{\nu\eta}\Gamma^\eta_{\mu\kappa}. \tag{4}$$

The utility of this tensor is that a given spacetime will be flat if and only if every component of $R^\lambda{}_{\mu\nu\kappa}$ is zero. When all components of $R^\lambda{}_{\mu\nu\kappa}$ are zero, (3) describes Newton's second law of motion for a free particle in the absence of gravity as viewed in an accelerating coordinate system. And when $R^\lambda{}_{\mu\nu\kappa}$ is non-zero, there is a choice of values for the Christoffel symbols (viz. the Schwarzschild metric ones that are associated with the vanishing of the Ricci tensor $R_{\mu\kappa}=R^\lambda{}_{\mu\lambda\kappa}$) that enables (3) to describe Newton's law of gravity, again in an arbitrary accelerating coordinate system. Then, with the Schwarzschild metric also giving relativistic corrections to Newtonian gravity, the observation of the predicted gravitational bending of light by the Sun established the validity of the above description of nature.

One can thus say with confidence that gravity is a covariant metric theory in which the metric describes the gravitational field, and that the geometry in the vicinity of the Sun is given by the Schwarzschild metric

$$ds^2=-B(r)dt^2+A(r)dr^2+r^2d\theta^2+r^2\sin^2\theta d\phi^2, \tag{5}$$

where

$$B(r)=A^{-1}(r)=1-2\beta/r,\qquad \beta=MG/c^2, \tag{6}$$

up to the perturbative order to which the metric has so far actually been tested. Thus any viable theory of gravity must embody all of the above, and we note that in the above we have not specified the equation of motion that is to be obeyed by the gravitational field. Rather, we have indicated only what its solution on solar system distance scales needs to look like.

2. EINSTEIN GRAVITY: WHAT COULD BE CHANGED

To complete the theory one thus needs to specify the gravitational field equations themselves. To this end Einstein postulated that the needed equations are to be of the form

$$-\frac{1}{8\pi G}\left(R^{\mu\nu}-\frac{1}{2}g^{\mu\nu}R^\alpha{}_\alpha\right)=T^{\mu\nu}_{\rm M}, \tag{7}$$

a set of equations that can be obtained via functional variation with respect to the metric of an action for the Universe of the form

$$I_{\rm UNIV} = I_{\rm EH} + I_{\rm M} = -\frac{1}{16\pi G}\int d^4x(-g)^{1/2}R^{\alpha}{}_{\alpha} + I_{\rm M}. \tag{8}$$

In (7) and (8) M denotes the matter field sector and $T^{\mu\nu}_{\rm M}$ is the matter field energy-momentum tensor. Given (7), we immediately see that in the source-free region where $T^{\mu\nu}_{\rm M} = 0$ the Ricci tensor has to vanish, with the solution given in (5) and (6) then following in the region exterior to a static, spherically symmetric source.

However, while (5) and (6) follow from (7), this is not the only way to secure the Ricci-flat Schwarzschild solution. Consider for example an action for the Universe of the conformal form

$$\begin{aligned} I_{\rm UNIV} = I_{\rm W} + I_{\rm M} &= -\alpha_g \int d^4x(-g)^{1/2} C_{\lambda\mu\nu\kappa}C^{\lambda\mu\nu\kappa} + I_{\rm M} \\ &= -2\alpha_g \int d^4x(-g)^{1/2}\left[R^{\mu\nu}R_{\mu\nu} - \frac{1}{3}(R^{\alpha}{}_{\alpha})^2\right] + I_{\rm M}, \end{aligned} \tag{9}$$

where

$$\begin{aligned} C_{\lambda\mu\nu\kappa} &= R_{\lambda\mu\nu\kappa} + \frac{1}{6}R^{\alpha}{}_{\alpha}\left[g_{\lambda\nu}g_{\mu\kappa} - g_{\lambda\kappa}g_{\mu\nu}\right] \\ &- \frac{1}{2}\left[g_{\lambda\nu}R_{\mu\kappa} - g_{\lambda\kappa}R_{\mu\nu} - g_{\mu\nu}R_{\lambda\kappa} + g_{\mu\kappa}R_{\lambda\nu}\right] \end{aligned} \tag{10}$$

is the Weyl conformal tensor. Functional variation of this action with respect to the metric leads to the equation of motion (see e.g. [1])

$$-4\alpha_g W^{\mu\nu} + T^{\mu\nu}_{\rm M} = 0, \tag{11}$$

where

$$\begin{aligned} W^{\mu\nu} &= \frac{1}{2}g^{\mu\nu}(R^{\alpha}{}_{\alpha})^{;\beta}{}_{;\beta} + R^{\mu\nu;\beta}{}_{;\beta} - R^{\mu\beta;\nu}{}_{;\beta} - R^{\nu\beta;\mu}{}_{;\beta} - 2R^{\mu\beta}R^{\nu}{}_{\beta} + \frac{1}{2}g^{\mu\nu}R_{\alpha\beta}R^{\alpha\beta} \\ &- \frac{2}{3}g^{\mu\nu}(R^{\alpha}{}_{\alpha})^{;\beta}{}_{;\beta} + \frac{2}{3}(R^{\alpha}{}_{\alpha})^{;\mu;\nu} + \frac{2}{3}R^{\alpha}{}_{\alpha}R^{\mu\nu} - \frac{1}{6}g^{\mu\nu}(R^{\alpha}{}_{\alpha})^2, \end{aligned} \tag{12}$$

to thus yield a gravitational theory that also has $R_{\mu\nu} = 0$ as a vacuum solution. This analysis thus shows that the Einstein equations given in (7) are only sufficient to give the Schwarzschild solution and its non-relativistic Newtonian limit but not necessary.

The Einstein equations are thus not uniquely selected. Moreover, in and of itself, the requirement that the gravitational action be a general coordinate scalar does not at all restrict the gravitational sector of the action to be of the Einstein-Hilbert form $I_{\rm EH}$ form given in (8), with the number of possible general coordinate invariant gravitational actions that one could write down actually being infinite, since one could use arbitrarily high powers of the Riemann tensor and its contractions. This lack of uniqueness is familiarly reflected in the fact that one is free to augment (8) with a term of the form $-\int d^4x(-g)^{1/2}\Lambda$ where Λ is the cosmological constant.

Beyond this we note that if one takes (7) as a given and extrapolates it beyond its weak classical gravity solar system origins, one runs into difficulties in essentially every type of extrapolation that one could consider. Thus if one extrapolates the classical theory to galaxies and clusters of galaxies one runs into the dark matter problem, if one extrapolates the classical theory to cosmology one runs into the cosmological constant or dark energy problem, if one extrapolates to strong classical gravity one runs into the singularity problem, and if one quantizes the theory and extrapolates far off the mass shell one runs into the renormalizability problem. Now even though no dark matter has yet been detected and dark energy is not yet at all understood, if one nonetheless takes dark matter and dark energy as a given, one then encounters many successes as well (such as big bang nucleosynthesis, anisotropy of the cosmic microwave background, strong lensing). However, to achieve these successes one has to take the energy budget of the Universe to be of order 70% dark energy and 25% dark matter, with only 5% or so being regular luminous baryonic matter. Not only is there as yet no detection of the needed dark matter particles, the required amount of dark energy is 60 orders of magnitude or so less than the amount expected from fundamental elementary particle physics – and if one were to use the large particle physics value the fits would be disastrous. Moreover, with all applications to date of gravity to astrophysics and cosmology having been made with gravity itself being treated classically, there is no guarantee that the current successes of standard Einstein gravity would not be modified by its non-renormalizable quantum corrections. Given these concerns we shall thus look for a completely different extrapolation of solar system wisdom, in a theory that is unambiguously specified. As we shall see, via the imposition of a particular invariance principle, namely local conformal invariance, none of the above extrapolation problems or ambiguities will any longer be encountered.

3. CONFORMAL GRAVITY: AN AB INITIO APPROACH

If we start with the kinetic energy of a free massless fermion in flat spacetime, then in order to obtain an action that is locally gauge invariant under $\psi(x) \to e^{i\beta(x)}\psi(x)$ (we suppress internal symmetry group indices), we introduce a gauge field that transforms as $A_\mu(x) \to A_\mu(x) + \partial_\mu\beta(x)$, and minimally couple according to

$$I_{\rm M} = -\int d^4x\bar{\psi}(x)\gamma^\mu[i\partial_\mu + A_\mu(x)]\psi(x). \tag{13}$$

Similarly, if we start with the kinetic energy of a free massless fermion in flat spacetime, in order to obtain an action that is locally coordinate invariant, we introduce the fermion spin connection $\Gamma_\mu(x)$, with the action taking the form

$$I_{\rm M} = -\int d^4x(-g)^{1/2}\bar{\psi}(x)\gamma^\mu(x)[i\partial_\mu + i\Gamma_\mu(x)]\psi(x), \tag{14}$$

where $\gamma^\mu(x) = V_a^\mu(x)\hat{\gamma}^a$ and $\Gamma_\mu(x) = [\gamma^\nu(x), \partial_\mu\gamma_\nu(x)]/8 - [\gamma^\nu(x), \gamma_\sigma(x)]\Gamma^\sigma_{\mu\nu}/8$, with $V_a^\mu(x)$ being a vierbein and the four $\hat{\gamma}^a$ being the special-relativistic Dirac gamma matrices.

Having now obtained (14) this way, we find that without having required it explicitly, this action actually has an additional symmetry, as it is locally conformal invariant under $\psi(x) \to e^{-3\alpha(x)/2}\psi(x)$, $V^a_\mu(x) \to e^{\alpha(x)}V^a_\mu(x)$, $g_{\mu\nu}(x) \to e^{2\alpha(x)}g_{\mu\nu}(x)$. Consequently, we can regard the spin connection $\Gamma_\mu(x)$ as being introduced not to maintain local coordinate invariance but rather to maintain local conformal invariance instead, in exactly the same minimally-coupled way that $A_\mu(x)$ maintains local gauge invariance. Thus if we require that the kinetic energy of massless fermion be invariant under complex phase transformations of the form $\psi(x) \to e^{-3\alpha(x)/2+i\beta(x)}\psi(x)$, we will be led to an action

$$I_{\rm M} = -\int d^4x(-g)^{1/2}\bar{\psi}(x)\gamma^\mu(x)[i\partial_\mu + i\Gamma_\mu(x) + A_\mu(x)]\psi(x) \tag{15}$$

that is locally gauge invariant and locally conformal invariant combined. (Under a local gauge transformation the metric transforms as $g_{\mu\nu}(x) \to g_{\mu\nu}(x)$, while under a local conformal transformation the gauge field transforms as $A_\mu(x) \to A_\mu(x)$.)

The reason why such a local conformal structure emerges in (14) is that massless particles move on the light cone, and the light cone is not just Poincare invariant, it is invariant under the full 15-parameter conformal group $O(4,2)$. (If the $ds^2 = g_{\mu\nu}(x)dx^\mu dx^\nu$ line element is zero, then so is $ds^2 = e^{2\alpha(x)}g_{\mu\nu}(x)dx^\mu dx^\nu$.) Moreover, the covering group of $O(4,2)$ is $SU(2,2)$. Since $SU(2,2)$ is generated by the 15 Dirac matrices (γ_5, γ_μ, $\gamma_\mu\gamma_5$, $[\gamma_\mu,\gamma_\nu]$), its fundamental representation is a fermionic field, and the full conformal structure of the light cone is thus built into a massless fermionic field.

As we see, it is thus natural to take fermions to be the most basic elements in physics, with internal symmetry gauge fields and a gravitational spin connection being induced when one gives the fermion kinetic energy a local complex phase invariance. However, such a starting point does not generate any kinetic energy terms for the gauge and gravitational fields. To actually generate them rather than just postulate them we take note of a calculation by 't Hooft [2]. Specifically, 't Hooft evaluated the logarithmically divergent part of the path integral $\int D\bar{\psi}D\psi\exp(iI_{\rm M})$ associated with (15), and found that after dimensional regularization it took the form of an effective action:

$$I_{\rm EFF} = \int d^4x(-g)^{1/2}C\left[\frac{1}{20}[R^{\mu\nu}R_{\mu\nu} - \frac{1}{3}(R^\alpha{}_\alpha)^2] + \frac{1}{3}F_{\mu\nu}F^{\mu\nu}\right], \tag{16}$$

where $C = 1/8\pi^2(4-D)$ in spacetime dimension D. Comparing with (9) we see that we have generated none other than the conformal gravity action $\int d^4x(-g)^{1/2}C_{\lambda\mu\nu\kappa}C^{\lambda\mu\nu\kappa}$ (as then rewritten using the Gauss-Bonnet theorem) together with the Maxwell action. With the Maxwell action being invariant under $A_\mu(x) \to A_\mu(x) + \partial_\mu\beta(x)$ and with the conformal gravity action being invariant under $g_{\mu\nu}(x) \to e^{2\alpha(x)}g_{\mu\nu}(x)$ (see e.g. [1]), the conformal gravity action thus serves as the gravitational analog of the Maxwell action, and in the following we shall thus use local conformal invariance as the principle with which to fix the structure of the gravitational action. In so doing we see that gravity can be generated by gauging the full conformal symmetry of the light cone.

Given the assumption of local conformal invariance, we find that the action $I_{\rm W} = -\alpha_g\int d^4x(-g)^{1/2}C_{\lambda\mu\nu\kappa}C^{\lambda\mu\nu\kappa}$ is the unique gravitational action that is invariant under the local transformation $g_{\mu\nu}(x) \to e^{2\alpha(x)}g_{\mu\nu}(x)$, with the gravitational

coupling constant α_g being dimensionless. Because α_g is dimensionless the conformal theory is power-counting renormalizable, and thus if we did not include an initial $\int d^4x(-g)^{1/2}C_{\lambda\mu\nu\kappa}C^{\lambda\mu\nu\kappa}$ term in the action, we would anyway generate one as a renormalization counter-term. Because the variation of the conformal action leads to fourth-order equations of motion, it had long been thought that the theory would not be unitary. However, as we describe in Sec. (5), Bender and Mannheim [3, 4] have recently shown that one can find a realization of the theory that is unitary. Consequently, conformal gravity is to be regarded as a bona fide quantum gravitational theory. Moreover, the similarity of the theory to Maxwell theory also carries over to the generation of a macroscopic classical limit starting from a microscopic quantum field theory. Specifically, in just the same way as the classical Maxwell equations emerge from the quantum Maxwell equations as matrix elements of the quantum fields in states containing an indefinite number of photons, because of its renormalizability the same will happen for conformal gravity in states containing an indefinite number of gravitational quanta, with both the quantum theory and its macroscopic classical limit obeying the equation of motion given in (11). Consequently, in the following we will be able to use (11) to study both the microscopic zero-point fluctuation problem and the macroscopic behavior of the theory on astrophysical distance scales.

The non-renormalizable Einstein-Hilbert action is expressly forbidden by the conformal symmetry because Newton's constant carries an intrinsic dimension. However, as noted above, this does not prevent the theory from possessing the Schwarzschild solution and Newton's law of gravity. In addition, the same conformal symmetry forbids the presence of any intrinsic cosmological constant term as it carries an intrinsic dimension too; with conformal invariance thus providing a very good starting point for tackling the cosmological constant problem.

Now we recall that the fermion and gauge boson sector of the standard $SU(3) \times SU(2) \times U(1)$ model of strong, electromagnetic, and weak interactions is also locally conformal invariant since all the associated coupling constants are dimensionless, and gauge bosons and fermions get masses dynamically via spontaneous symmetry breaking. Other than the Higgs sector (which we shall shortly dispense with), the standard model Lagrangian is devoid of any intrinsic mass or length scales. And with its associated energy-momentum tensor serving as the source of gravity, it is thus quite natural that gravity should be devoid of any intrinsic mass or length scales too. Our use of conformal gravity thus nicely dovetails with the standard $SU(3) \times SU(2) \times U(1)$ model. To tighten the connection, we note that while the standard $SU(3) \times SU(2) \times U(1)$ model is based on second-order equations of motion, an electrodynamics Lagrangian of the form $F^{\mu\nu}\partial_\alpha\partial^\alpha F_{\mu\nu}$ would be just as gauge and Lorentz invariant as the Maxwell action, and there is no immediate reason to leave any such type of term out. Now while an $F^{\mu\nu}\partial_\alpha\partial^\alpha F_{\mu\nu}$ theory would not be renormalizable, in and of itself renormalizability is not a law of nature (witness Einstein gravity). However, such a theory would not be conformal invariant. Thus if we impose local conformal invariance as a principle, we would then force the fundamental gauge theories to be second order, and thus be renormalizable after all. However, imposing the same symmetry on gravity expressly forces it to be fourth order instead, with gravity then also being renormalizable. As we see, renormalizability is thus a consequence of conformal invariance.

Now if the underlying theory is to be locally conformal invariant, there would be no place for a fundamental Higgs field with its tachyonic double-well potential. Instead, mass scales would have to be generated dynamically in the vacuum via dynamical fermion bilinear condensates. The elimination of a fundamental Higgs field has an immediate benefit – one no longer has to deal with the uncontrollable cosmological constant contribution it produces when it acquires a non-vanishing expectation value. However, one still has to make contact with the standard model, and one would thus want to obtain a standard model Lagrangian with some effective scalar field. Such an effective scalar field would have to emerge as a Ginzburg-Landau c-number order parameter, i.e. as the matrix element of a fermion bilinear operator in some possibly spacetime-dependent coherent state. Such an effective c-number scalar field would not be observable in an accelerator.

To see what such an effective theory might look like, we note that if the fermion acquires a mass parameter $M(x)$ by some dynamical symmetry breaking mechanism, the associated Hartree-Fock mean-field action would take the same form as given in (15), only with a mass term added, viz.

$$I_{\rm M} = -\int d^4x(-g)^{1/2}\bar{\psi}(x)\gamma^\mu(x)[i\partial_\mu + i\Gamma_\mu(x) + A_\mu(x) + M(x)]\psi(x). \qquad (17)$$

Evaluating the divergent part of the same $\int D\bar{\psi}D\psi\exp(iI_{\rm M})$ path integral as before only with (17) this time generates the same $I_{\rm EFF}$ as in (16), while adding on the mean-field action [2] (as written here using the sign convention employed in this paper for $R^\alpha{}_\alpha$):

$$I_{\rm MF} = \int d^4x(-g)^{1/2}C\left[-M^4(x) + \frac{1}{6}M^2(x)R^\alpha{}_\alpha - g_{\mu\nu}\partial^\mu M(x)\partial^\nu M(x)\right]. \qquad (18)$$

Here C is the same logarithmically divergent constant as before. Finally, if we give $M(x)$ a group index, the same procedure would cause the $\partial^\mu M(x)$ terms to be replaced by covariant gauge derivatives (see [5, 6]), and would yield

$$\begin{aligned} I_{\rm MF} &= \int d^4x(-g)^{1/2}C\Big[-M^4(x) + \frac{1}{6}M^2(x)R^\alpha{}_\alpha \\ &- g_{\mu\nu}(\partial^\mu + iA^\mu(x))M(x)(\partial^\nu - iA^\nu(x))M(x)\Big]. \end{aligned} \qquad (19)$$

When (19) is taken in conjunction with (16), a conformally coupled standard model emerges, but with a c-number scalar parameter $M(x)$ that is not a fundamental field. In Secs. (7) and (8) we will show how such a mass parameter $M(x)$ could be generated as a fermion bilinear condensate in a conformal invariant theory. And in [7] we explore whether or not the existence of such a c-number order parameter $M(x)$ necessitates the existence of an accompanying dynamical bound state scalar particle.

4. QUANTIZATION OF GRAVITY THROUGH COUPLING

In trying to find solutions to the conformal theory, we note that given the lack of any intrinsic mass or length scales in the conformal action, without dynamical generation

of such scales there could be no non-trivial solutions to the theory. Thus if all mass and length scales are to come from dynamics, and all such mass-generating dynamics is to be quantum-mechanical (c.f. no fundamental Higgs fields), the only allowed geometry in purely classical conformal gravity would be one with no curvature at all, viz. a geometry that is Minkowski. (For there to be any curvature one needs some length scale to characterize it.) Thus if one takes $W^{\mu\nu}$ to be a classical fourth-order derivative function, then even though one could readily find an exterior vacuum solution to $W^{\mu\nu} = 0$ such as the Ricci-flat one given in (5) and (6), there would be no basis for taking the dimensionful parameter β to be non-zero since the theory is as yet scale free. For the solution to a differential equation to have lower symmetry than the equation itself it (i.e. for a solution to contain symmetry-breaking integration constants that are not present in the equation itself), there has to be a spontaneous breakdown of the symmetry. If all symmetry breaking is to be quantum-mechanical, then in the absence of quantum mechanics, geometry would have to be flat. Thus, if the mass that appears in the Schwarzschild radius of a source is quantum-mechanically generated, then the curvature produced by that mass is ipso facto due to quantum mechanics too.

The above remarks require some clarification since in classical electrodynamics one can construct plane wave solutions with $k^0 = |\bar{k}|$ even though the classical Maxwell equations are conformal invariant. Specifically, if we have a free classical Maxwell action $\int d^4x F_{\mu\nu}F^{\mu\nu}$ in flat spacetime, the associated homogeneous wave equation takes the form $(\partial_t^2 - \nabla^2)A_\mu(x) = 0$ (in the convenient Lorentz gauge). However, while this equation has $A_\mu(x) = \varepsilon_\mu(x)\exp(-ik\cdot x)$ as a solution, this solution involves a dimensionful four-vector momentum k_μ that is not present in the equation of motion itself. Hence some mechanism is required to generate such a four-momentum. In classical Maxwell theory the mechanism for doing this is not from the $F_{\mu\nu}$ sector of the theory at all, but rather via the introduction of a localized source $J_\mu(x)$. In the presence of the source the solution to the inhomogeneous $(\partial_t^2 - \nabla^2)A_\mu(x) = J_\mu(x)$ is given as $A_\mu(x) = \int d^4x' D(x-x')J_\mu(x')$ where $D(x-x')$ is the massless retarded propagator. Now $D(x) = \delta(t-r)/4\pi r$ itself is written entirely in terms of the spacetime coordinates and contains no fundamental scales. Rather, the scales reside in the localized $J_\mu(x)$, and if it for instance oscillates with a specific frequency, then the resulting $A_\mu(x)$ will oscillate with that same frequency too. However, for $J_\mu(x)$ to be localized and possess such an oscillation frequency, it would not be scale invariant. Hence the classical Maxwell field only possesses frequency scales because the sources to which it couples are taken to possess them, and the sources themselves can only possess such scales if they are not scale invariant. However, if all the particles contained in electromagnetic sources are to acquire length scales quantum-mechanically, then there could be no fundamental classical electromagnetic sources that could possess such length scales in the first place. Classical electromagnetism with localized oscillating sources is thus a macroscopic manifestation of an underlying microscopic quantum Maxwell theory in which scales are generated dynamically. In a truly scale-free classical electrodynamics there would be no propagation of electromagnetic waves. Thus just like gravity, the same conformal invariance will not permit electromagnetic sources to have any nontrivial intrinsically classical component either, with such sources being intrinsically quantum-mechanical.

If momentum modes are not to arise in classical physics, one needs to ask how

it is that they then do arise. To this end we note that as well as generating mass scales via dynamics, there is another way in which quantum mechanics produces scales, namely via the quantization procedure itself. Specifically, scales are introduced via canonical commutation relations, with the generic equal-time commutation relation $[\phi(\bar{x},t),\pi(\vec{x}',t)] = i\hbar\delta^3(\bar{x}-\vec{x}')$ for instance being a non-linear relation that introduces a scale $\delta^3(\bar{x}-\vec{x}')$ everywhere on a spacelike hypersurface. Since we can set $\delta^3(\bar{x}-\vec{x}') = (1/8\pi^3)\int d^3k \exp(i\bar{k}\cdot(\bar{x}-\vec{x}'))$, this is equivalent to introducing a complete basis of momentum modes, with momentum modes thus being intrinsically quantum mechanical. And indeed it is the very existence of this set of modes that gives rise to the zero-point energy density and pressure of a quantized field that we discuss in the following. Moreover, a quantized field will have a zero-point energy density and pressure even when it is massless, i.e. even in the absence of mass generation. Then, when there is mass generation, the momentum modes will obey the $k_0^2 = \bar{k}^2 + m^2/\hbar^2$ mass condition and cause the massless theory zero-point energy density and pressure to readjust. And as we show in Secs. (7) and (8), this readjustment will cancel the cosmological constant that is induced by the same mass generation mechanism.

Given the above remarks, we see that in conformal gravity we should expand the metric as a power series in Planck's constant rather than as a power series in the gravitational coupling constant, with the zeroth-order term in the expansion being flat. Consequently, in the theory there is no intrinsic classical gravity, with the equations that are to be used for macroscopic systems being associated with matrix elements of the quantum fields in states with an indefinite number of gravitational quanta, Since there is to be no intrinsic classical gravity, there could not be any classical black holes. While conformal gravity thus eliminates the classical gravity singularity problem, and thus simultaneously eliminates the need to have to make such classical singularities compatible with quantum mechanics, it remains to be seen whether the theory might still generate geometric singularities through quantum-mechanical effects, though one might anticipate that the uncertainty principle might spread sources out enough to prevent this from happening.

Even with the requirement that the metric be expanded as a power series in Planck's constant, quantization of gravity can still not follow the standard canonical quantization prescription that is used for other fields. Specifically, for a matter field one obtains its equation of motion by varying the matter action with respect to the matter field, but one obtains its energy-momentum tensor $T_M^{\mu\nu}$ by instead varying the matter action with respect to the metric. Since $T_M^{\mu\nu}$ involves products of matter fields at the same point, a canonical quantization of the matter field then gives the matter energy-momentum tensor a non-vanishing zero-point contribution. However, in a standard quantization procedure for a given matter field, the non-vanishing of $T_M^{\mu\nu}$ violates no constraint since one does not simultaneously impose the equation of motion of any other field. Thus for a given matter field one does not require stationarity with respect to the metric, with $T_M^{\mu\nu}$ thus not being constrained to vanish.

In contrast however, for gravity the relevant field is the metric itself. If we define the variation of the gravitational action with respect to the metric to be a quantity $T_{GRAV}^{\mu\nu}$, the gravitational equation of motion is then given by $T_{GRAV}^{\mu\nu} = 0$. Then, with $T_{GRAV}^{\mu\nu}$ containing products of fields at the same point, a canonical quantization of the gravitational field

would give a zero-point contribution to $T^{\mu\nu}_{GRAV}$, and thus violate the stationarity condition $T^{\mu\nu}_{GRAV} = 0$ that $T^{\mu\nu}_{GRAV}$ has to obey. Hence, unlike the matter fields for which there is no constraint on $T^{\mu\nu}_{M}$ in the absence of any coupling of matter to gravity, gravity itself is always coupled to gravity, with its stationarity condition not permitting it to consistently be quantized on its own.

Despite this, we note that if we impose a stationarity condition with respect to the metric not on the gravity piece or the matter piece alone, but on their sum as given by the total $I_{\rm UNIV}$ of the universe introduced in (9), we then obtain

$$T^{\mu\nu}_{\rm UNIV} = T^{\mu\nu}_{\rm GRAV} + T^{\mu\nu}_{\rm M} = 0. \tag{20}$$

In this case it now is possible to quantize gravity consistently, with $T^{\mu\nu}_{\rm GRAV}$ now being able to be non-zero provided gravity is coupled to some quantized matter field source for which $T^{\mu\nu}_{\rm M}$ is non-zero. Thus gravity can only be quantized consistently if it is coupled to a quantized matter field. However, in order for the cancellation required of the total $T^{\mu\nu}_{\rm UNIV}$ to actually take place, the quantization condition imposed on the gravitational sector commutation relations would have to be fixed by the quantization condition in the matter sector in order to enforce $T^{\mu\nu}_{\rm GRAV} = -T^{\mu\nu}_{\rm M}$, with each term being intrinsically quantum-mechanical. Consequently, gravity is not only quantized though its coupling to quantized matter, its commutation relations are explicitly determined by that coupling, with gravity needing no independent quantization of its own. Finally, we note that not only do the matter fields quantize gravity, the vanishing of $T^{\mu\nu}_{\rm UNIV}$ entails that the gravity field and the matter field zero-point fluctuations must cancel each other identically. In Secs. (6), (7) and (8) we explore this point in detail.

5. UNITARITY VIA *PT* SYMMETRY

When the $W^{\mu\nu}$ tensor given in (12) is linearized around a flat spacetime background with metric $\eta_{\mu\nu}$ according to $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$, it is found [7] to be a function of the traceless quantity $K^{\mu\nu} = h^{\mu\nu} - (1/4)\eta^{\mu\nu}\eta_{\alpha\beta}h^{\alpha\beta}$. In the convenient transverse gauge $\partial_\mu K^{\mu\nu} = 0$, the first order term in $W^{\mu\nu}$ is found to take the simple form

$$W^{\mu\nu}(1) = \frac{1}{2}(\partial_\alpha\partial^\alpha)^2 K^{\mu\nu}, \tag{21}$$

while the second order term in the conformal action $I_{\rm W}$ given in (9) takes the form

$$I_{\rm W}(2) = -\frac{\alpha_g}{2}\int d^4x \partial_\alpha\partial^\alpha K_{\mu\nu}\partial_\beta\partial^\beta K^{\mu\nu}. \tag{22}$$

Since there is no mixing of components of $K^{\mu\nu}$ in either (21) or (22), one can explore the unitarity structure of the theory by working with an analog one-component scalar field theory. As such, the condition $W^{\mu\nu}(1) = 0$ is one of a broad class of fourth-order equations of motion that have been encountered in the literature, and all of them can be associated with the generic scalar action

$$I_{\rm S} = -\frac{1}{2}\int d^4x \left[\partial_\mu\partial_\nu\phi\partial^\mu\partial^\nu\phi + (M_1^2 + M_2^2)\partial_\mu\phi\partial^\mu\phi + M_1^2M_2^2\phi^2\right]. \tag{23}$$

Given this action one obtains an equation of motion

$$(-\partial_t^2+\bar{\nabla}^2-M_1^2)(-\partial_t^2+\bar{\nabla}^2-M_2^2)\phi(x)=0, \tag{24}$$

a propagator

$$D(k,M_1,M_2)=\frac{1}{(M_2^2-M_1^2)}\left(\frac{1}{k^2+M_1^2}-\frac{1}{k^2+M_2^2}\right), \tag{25}$$

and an energy-momentum tensor with $(0,0)$ component

$$T_{00}(M_1,M_2)=\pi_0\dot{\phi}+\frac{1}{2}\left[\pi_{00}^2+(M_1^2+M_2^2)(\dot{\phi}^2-\partial_i\phi\partial^i\phi)-M_1^2M_2^2\phi^2-\pi_{ij}\pi^{ij}\right], \tag{26}$$

where

$$\begin{aligned}\pi^\mu &= \frac{\partial\mathscr{L}}{\partial\phi_{,\mu}}-\partial_\lambda\left(\frac{\partial\mathscr{L}}{\partial\phi_{,\mu,\lambda}}\right)=-(M_1^2+M_2^2)\partial^\mu\phi+\partial_\lambda\partial^\mu\partial^\lambda\phi,\\ \pi^{\mu\lambda} &= \frac{\partial\mathscr{L}}{\partial\phi_{,\mu,\lambda}}=-\partial^\mu\partial^\lambda\phi.\end{aligned} \tag{27}$$

These equations immediately possess two well-known realizations that exhibit the problems that higher-derivative theories are thought to possess. If one takes the contour for the k^0 integration in the propagator to be the standard Feynman one in which all positive energy modes propagate forward in time and all negative energy modes propagate backwards in time, because of the relative minus sign in (25), one finds that some of the poles have negative residues, with the theory being quantized with an indefinite metric. To avoid such negative residues, one can find [4] an alternate contour in which the residues of all poles are positive but in which some of the negative energy modes propagate forward in time. While one can quantize this realization with a standard Dirac norm, as the presence of the $-M_1^2M_2^2\phi^2$ term in $T_{00}(M_1,M_2)$ indicates, in this case the energy eigenvalue spectrum is unbounded from below.

With neither of these two possibilities being palatable, higher-derivative theories have long been regarded as being unphysical. However, recently Bender and Mannheim revisited the issue [3, 4] and found a third realization of the theory in which the energy spectrum is bounded from below and there are no negative Hilbert space norms at all. With the appropriate scalar product for a Hilbert space being determined by boundary conditions, to determine the relevant scalar product one needs some asymptotic information. To this end Bender and Mannheim studied the eigenvalue problem for the Hamiltonian $H=\int d^3x T_{00}(M_1,M_2)$ in the sector of the theory where the energy eigenvalue spectrum is bounded from below, and found that the associated wave functions were not normalizable on the real axis. In consequence of this, the Hamiltonian of the system could not be Hermitian. However, the wave functions were found to be normalizable on the imaginary axis, and thus the field ϕ would have to be anti-Hermitian rather than Hermitian, with the $-M_1^2M_2^2\phi^2$ term in $T_{00}(M_1,M_2)$ then being bounded from below. In addition they noted that if they constructed a path integral for the system, it would not exist with real ϕ but would be well-defined if ϕ were pure imaginary. So again, one needs to take ϕ to be an anti-Hermitian operator.

Now if a Hamiltonian is not Hermitian, one is immediately concerned that its eigenvalues might not all be real. However, while Hermiticity implies reality of eigenvalues, there is no converse theorem that says that a non-Hermitian Hamiltonian must have complex eigenvalues. Consequently, Hermiticity is only sufficient for reality but not necessary. Recently, as part of the general PT symmetry program that has been developed by Bender and collaborators [8] a necessary condition for reality has been found, namely that a Hamiltonian have a PT symmetry where P is a linear operator and T is an antilinear one. Specifically, it was shown in [9] that if a Hamiltonian is PT invariant the secular equation $|H-\lambda I|=0$ that determines the eigenvalues is real. Then in [10] the converse was shown, namely that if the secular equation is real, the Hamiltonian must have a PT symmetry. Consequently, the energy eigenspectrum of a Hamiltonian that is not PT symmetric must contain some complex eigenvalues.

Noting now that all the poles in the propagator (25) lie on the real axis, we see that the Hamiltonian for the fourth-order theory while not Hermitian must instead be PT symmetric. For such Hamiltonians one can construct a norm, the so-called PT norm [8], that obeys unitary time evolution. For our purposes here we note that for a non-Hermitian Hamiltonian H that has a completely real energy eigenspectrum, H and $H^\dagger$ must be related by a similarity transform of the form

$$VHV^{-1}=H^\dagger. \tag{28}$$

Thus if H has a right-eigenvector according to $H|R\rangle=E|R\rangle$ with real E, its conjugate will obey $\langle R|H^\dagger=\langle R|E$ and will not be a left-eigenvector of H. Rather, the state $\langle L|$ defined as $\langle L|=\langle R|V$ will be a left-eigenvector of H since it obeys

$$\langle L|H=\langle L|E. \tag{29}$$

In this case it will be the norm $\langle L|R\rangle=\langle R|V|R\rangle$ that will obey unitary time evolution since it evolves as

$$\langle L(t)|R(t)\rangle=\langle L(t=0)|e^{iHt}e^{-iHt}|R(t=0)\rangle=\langle L(t=0)|R(t=0)\rangle. \tag{30}$$

Thus we see that the needed scalar product is the overlap of a left-eigenvector with a right-eigenvector and not the overlap of a right-eigenvector with its own conjugate. Moreover, in [11] it was shown that the existence of a V that can connect H and $H^\dagger$ according to $VHV^{-1}=H^\dagger$ is a necessary and sufficient condition for both the existence of a PT symmetry and for the existence of a unitary scalar product, with the scalar product then being of the form $\langle L|R\rangle=\langle R|V|R\rangle$. The existence of a PT invariance for a Hamiltonian is thus a necessary and sufficient condition for unitary time evolution.

When these PT ideas are applied to the fourth-order propagator, it is found [4, 7] that the relative minus in it is no longer associated with an indefinite metric at all. Rather, it is associated with V operator, with the completeness relation being given by $\Sigma|n\rangle\langle n|V=I$ and not by $\Sigma|n\rangle\langle n|-\Sigma|m\rangle\langle m|=I$. Finally, the propagator itself is found to be given by the Green's function $\langle\Omega_L|T(\phi\phi)|\Omega_R\rangle=\langle\Omega_R|VT(\phi\phi)|\Omega_R\rangle$ rather than by the familiar $\langle\Omega_R|T(\phi\phi)|\Omega_R\rangle$. As we see, the unitarity problem for fourth-order propagators only arose because one wanted to represent them as $\langle\Omega_R|T(\phi\phi)|\Omega_R\rangle$. Once one recognizes that the left vacuum need not be the conjugate of the right vacuum unitarity can then readily be achieved.

Now while the above discussion was developed for the general second- plus fourth-order action given in (23), for the conformal case given in (21) we are interested in a pure fourth-order theory alone where the propagator is given by $D(k) = 1/k^4$ (a propagator whose poles again are all on the real axis). Since the reduction to a pure fourth-order theory would require setting both M_1^2 and M_2^2 equal to zero in (23), we see that because of the $1/(M_2^2 - M_1^2)$ prefactor in (25), the limit is singular. In consequence, the limit is a quite unusual one in which the Hamiltonian becomes a non-diagonalizable, Jordan-block Hamiltonian [4, 7], with some of the states that had been eigenstates being replaced by non-stationary ones. The very fact that the Hamiltonian is not diagonalizable immediately confirms that it could not be Hermitian, just as we had noted above. In this case even though the set of energy eigenstates is not complete, the set of stationary plus non-stationary states combined is complete [4], with time evolution of packets built out of the two classes of states combined being unitary [4]. The unitarity of the pure fourth-order conformal gravity theory is thus established.

As we thus see, in order to establish unitarity for fourth-order theories we need the field $\phi(x)$, and thus $g_{\mu\nu}(x)$ itself, to be anti-Hermitian rather than Hermitian. Now this is not how one ordinarily thinks about the gravitational field, since one would presuppose that it, above all fields, should have a real classical limit. Nonetheless, having an anti-Hermitian gravitational field is is not in conflict with anything that is actually known about gravity. Specifically, in [4] it was noted that if one replaces $g_{\mu\nu}$ by $ig_{\mu\nu}$ (and thus $g^{\mu\nu}$ by $-ig^{\mu\nu}$) neither the Christoffel symbols that appear in geodesics nor $R^{\lambda}{}_{\mu\nu\sigma}$ would be affected at all. In four space-time dimensions $\det(g_{\mu\nu})$ would not be affected either. Even though Riemann tensor contractions could generate factors of i, all such factors could be absorbed by redefining the overall multiplicative coefficients in the action (and likewise for the $ds^2 = g_{\mu\nu}dx^\mu dx^\nu$ line element). Hence, current gravitational measurements cannot distinguish between a purely real or a purely imaginary gravitational field. And as we have seen, once one takes the gravitational field to be anti-Hermitian, one can construct a consistent, renormalizable and unitary theory of quantum gravity. And perhaps the problems that beset quantum gravity have arisen because one wanted the gravitational field to be Hermitian.

6. THE ZERO-POINT PROBLEM

In current applications of standard gravity to macroscopic astrophysical and cosmological systems, one treats gravity itself as being purely classical. However, one cannot treat its matter source that way too since there are some intrinsically quantum-mechanical sources that are significant macroscopically. Thus, white dwarf stars are stabilized by the Pauli degeneracy pressure of the electrons in the star, and black-body radiation contributes to cosmic expansion. To couple these particular effects to classical gravity, both of them are taken to be described as ensemble averages over an appropriate set of positive-energy Fock space states. However, while the Fock space states would be eigenstates of a Hamiltonian of the generic form $\Sigma\hbar\omega(a^\dagger a + 1/2)$, the infinite $\Sigma\hbar\omega/2$ zero-point contribution is ignored, i.e. one takes the Hamiltonian to be of the truncated form $\Sigma\hbar\omega a^\dagger a$ instead. Now in flat space one is free to discard the zero-point term (say by

a normal ordering prescription) since in flat space one can only measure energy differences. However the hallmark of gravity is that it couple to energy itself and not to energy difference, and so discarding any contribution to the energy would require justification.

Since one would have to cancel infinities in the matter field energy-momentum tensor $T_{\rm M}^{\mu\nu}$ if the gravitational effects that occur in standard gravity are to be finite, some mechanism needs to be identified that would effect the cancellation. An immediate mechanism that might achieve this would be a cancellation between appropriately chosen matter fields, since bosons and fermions contribute to $T_{\rm M}^{\mu\nu}$ with opposite signs. In fact such a cancellation will occur if there is an exact supersymmetry between fermions and bosons. However, once the fermion-boson mass degeneracy is broken, the cancellation is lost. With the non-observation to date of any of the requisite superpartners, we know that the supersymmetry breaking scale must be at least in the TeV region, with the uncanceled zero-point energy then being huge.

Nonetheless, the generic idea of a boson-fermion cancellation as enforced by some symmetry still makes sense since fermions and bosons generate vacuum energies with opposite signs no matter what the theory. To take advantage of this, we note that the treatment of standard gravity described above is deficient in one rather serious regard, namely it ignores the effect of quantum mechanics on gravity itself. And as soon as one quantizes gravity, gravity itself will acquire a zero-point contribution. Since gravitational quanta are bosonic, under certain circumstances they may then be able to provide the needed cancellation. Specifically, for such a cancellation to occur one needs three things: a quantum gravity theory that makes sense, a symmetry, and the presence of fermions in the matter field sector. With conformal gravity meeting all of these criteria (as noted above fermions are its building blocks), in the following we will explore its implications for the vacuum energy problem. We will see that when the conformal symmetry is unbroken the needed cancellation does in fact occur. And then, unlike the supersymmetry situation, the cancellation will be maintained even after the conformal symmetry is spontaneously broken and a cosmological constant term is induced.

One of the challenges that gravity theory faces is that zero-point and cosmological constant contributions already occur for matter fields in flat spacetime, i.e. they occur in the absence of gravity. Since gravity is not involved in flat space physics, it is very difficult for gravity to then resolve any problem that it is not responsible for. To enable gravity to resolve these problems we need to put the gravitational field an equal footing with the matter fields. This we can do if there are no intrinsic classical contributions in either the gravitational or the matter sectors and all physics is quantum mechanical, i.e. precisely as conformal symmetry requires. Since the lowest order quantum-mechanical contribution to $T_{\rm M}^{\mu\nu}$ is a zero-point contribution of order $\hbar$, to cancel it through the vanishing of the total $T_{\rm UNIV}^{\mu\nu}$ given in (20), we will need the lowest non-trivial gravitational term to be of order $\hbar$ too. Since the zero-point contribution is due to products of fields at the same point, the order $\hbar$ gravitational zero-point must involve a product of two gravitational fields and thus be given by the second-order $-4\alpha_g W^{\mu\nu}(2)$ ($\equiv T_{\rm GRAV}^{\mu\nu}$) tensor that is obtained by varying the $I_{\rm W}(2)$ term in (22). To this order in $\hbar$ we only need to evaluate $T_{\rm M}^{\mu\nu}$ in a flat background. It will then generate an order $\hbar$ curvature, with the order $\hbar^2$ term in $T_{\rm M}^{\mu\nu}$ then being curvature dependent. Moreover, with $T_{\rm UNIV}^{\mu\nu}$ vanishing not only in lowest order but in all orders if both the gravitational and matter

field sectors are renormalizable (i.e. in a renormalizable theory (20) is an all-order identity), the zero-point cancellation will occur to all orders. Thus if we decompose $T^{\mu\nu}_{\rm GRAV}$ and $T^{\mu\nu}_{\rm M}$ into finite and divergent parts according to $T^{\mu\nu}_{\rm GRAV} = (T^{\mu\nu}_{\rm GRAV})_{\rm FIN} + (T^{\mu\nu}_{\rm GRAV})_{\rm DIV}$, $T^{\mu\nu}_{\rm M} = (T^{\mu\nu}_{\rm M})_{\rm FIN} + (T^{\mu\nu}_{\rm M})_{\rm DIV}$, (20) will decompose into

$$(T^{\mu\nu}_{\rm GRAV})_{\rm DIV} + (T^{\mu\nu}_{\rm M})_{\rm DIV} = 0, \tag{31}$$

and

$$(T^{\mu\nu}_{\rm GRAV})_{\rm FIN} + (T^{\mu\nu}_{\rm M})_{\rm FIN} = 0. \tag{32}$$

With (31) we see that all gravitational and matter field infinities cancel each other identically, with $(T^{\mu\nu}_{\rm GRAV})_{\rm DIV}$ and $(T^{\mu\nu}_{\rm M})_{\rm DIV}$ regulating each other. Given this regulation, there is no need to renormalize either of the two terms as their sum is finite, and thus no conformal anomaly is generated. (Alternatively, if one did first renormalize each term separately, the associated conformal anomalies would then cancel each other identically.) And with all infinities having been removed, (32) provides us with a completely finite framework for calculating gravitational effects. Thus in (31) we take care of the $\Sigma\hbar\omega/2$ type terms, and in (32) we are free to use the $\Sigma\hbar\omega a^{\dagger}a$ type terms alone.

With $W^{\mu\nu}(2)$ being of order $\hbar$, the gravitational fluctuation $K^{\mu\nu}$ must itself be of order $\hbar^{1/2}$, and since the lowest non-trivial term in $T^{\mu\nu}_{\rm M}$ is of order $\hbar$, it must be the case that $W^{\mu\nu}(1)$ in (21) vanish identically. While the vanishing of $W^{\mu\nu}(1)$ provides us with a wave equation, the situation is not quite the same as the one that occurs when one expands in a power series in the gravitational coupling constant. Specifically, in that case the first-order fluctuation term on the gravitational side is produced by a first-order fluctuation term on the matter side, so that the gravitational fluctuation would obey an inhomogeneous wave equation with a source. In contrast, in the conformal case the first-order gravitational wave equation is strictly homogeneous on all scales. Then, since this equation is homogeneous, in and of itself it does not force $K^{\mu\nu}$ to be non-zero. However, since the order $\hbar$ contribution to $T^{\mu\nu}_{\rm M}$ is non-zero, $-4\alpha_g W^{\mu\nu}(2)$ cannot vanish, and thus $K^{\mu\nu}$ cannot vanish either. It is thus quantization of the matter field that forces the gravitational field to be quantized, with the condition $-4\alpha_g W^{\mu\nu}(2) + T^{\mu\nu}_{\rm M} = 0$ fixing the strength of the commutator terms present in the second order $W^{\mu\nu}(2)$. With the matter field fixing the strength of the gravitational sector, the cancelation of both zero-point contributions and conformal anomalies is secured.

To see how things work in detail we consider conformal gravity coupled to a Dirac fermion. To the order $\hbar$ of interest to us we can take the fermion to be a free massless fermion in flat spacetime. In this case the matter field energy-momentum tensor is given by $T^{\mu\nu}_{\rm M} = i\hbar\bar{\psi}\gamma^{\mu}\partial^{\nu}\psi$, with its vacuum expectation value being given by

$$\langle\Omega|T^{\mu\nu}_{\rm M}|\Omega\rangle = -\frac{2\hbar}{(2\pi)^3}\int_{-\infty}^{\infty} d^3k \frac{k^{\mu}k^{\nu}}{\omega_k}, \tag{33}$$

where k^{μ} is a lightlike 4-vector $k^{\mu} = (\omega_k, \bar{k})$ with $\omega_k = |\bar{k}|$. In (33) we recognize two separate infinite terms, one associated with $\rho_{\rm M} = \langle\Omega|T^{00}_{\rm M}|\Omega\rangle$ and the other with $p_{\rm M} = \langle\Omega|T^{11}_{\rm M}|\Omega\rangle = \langle\Omega|T^{22}_{\rm M}|\Omega\rangle = \langle\Omega|T^{33}_{\rm M}|\Omega\rangle$. Since the fermion is massless, $T^{\mu\nu}_{\rm M}$ is traceless and thus these two infinities obey $\rho_{\rm M} = 3p_{\rm M}$. Such a form could not be

associated with a cosmological constant term of the form $-\Lambda\eta_{\mu\nu}$ since its trace is given by the non-zero -4Λ. Rather, given its $k^\mu k^\nu$ structure, the quantity $\langle\Omega|T_{\rm M}^{\mu\nu}|\Omega\rangle$ can be written in the form of a perfect fluid with a timelike fluid velocity vector $U^\mu = (1,0,0,0)$, viz.

$$\langle\Omega|T_{\rm M}^{\mu\nu}|\Omega\rangle = (\rho_{\rm M} + p_{\rm M})U^\mu U^\nu + p\eta^{\mu\nu}, \tag{34}$$

with the fluid thus possessing both a zero-point energy density and a zero-point pressure. Since gravity couples to the full $T_{\rm M}^{\mu\nu}$ and not just to its $(0,0)$ component, it is not sufficient to only address the vacuum energy density problem, one has to deal with the vacuum pressure as well. There are thus two vacuum problems that need to be addressed, and not just one. The gravitational sector will thus need to cancel both the vacuum energy density and the vacuum pressure of the matter field, and in a conformal theory will readily be able to do so since $-4\alpha_g W^{\mu\nu}(2)$ has an identical traceless vacuum perfect fluid structure. (In a conformal invariant theory the variation with respect to the metric of the pure gravitational sector of the action is automatically traceless.)

For the explicit structure of the gravity sector we follow the discussion given in [7]. On using some residual gauge symmetry the general solution to $W^{\mu\nu}(1) = 0$ is given as

$$\begin{aligned} K_{\mu\nu}(x) &= \frac{\hbar^{1/2}}{2(-\alpha_g)^{1/2}}\sum_{i=1}^{2}\int\frac{d^3k}{(2\pi)^{3/2}(\omega_k)^{3/2}}\Big[A^{(i)}(\overline{k})\varepsilon_{\mu\nu}^{(i)}(\overline{k})e^{ik\cdot x} \\ &+ i\omega_k B^{(i)}(\overline{k})\varepsilon_{\mu\nu}^{(i)}(\overline{k})(n\cdot x)e^{ik\cdot x} \\ &+ \hat{A}^{(i)}(\overline{k})\varepsilon_{\mu\nu}^{(i)}(\overline{k})e^{-ik\cdot x} - i\omega_k\hat{B}^{(i)}(\overline{k})\varepsilon_{\mu\nu}^{(i)}(\overline{k})(n\cdot x)e^{-ik\cdot x}\Big], \end{aligned} \tag{35}$$

as expressed in terms of quantum operators $A^{(i)}(\overline{k})$, $\hat{A}^{(i)}(\overline{k})$, $B^{(i)}(\overline{k})$ and $\hat{B}^{(i)}(\overline{k})$ and two transverse traceless polarization tensors $\varepsilon_{\mu\nu}^{(i)}(\overline{k})$ $(i = 1,2)$, both of which are normalized to $\varepsilon_{\alpha\beta}\varepsilon^{\alpha\beta} = 1$. Since $K^{\mu\nu}$ is to not be Hermitian, the creation operators are not the Hermitian conjugates of the annihilation operators. However, in the following all that will matter is that $A^{(i)}(\overline{k})$ and $B^{(i)}(\overline{k})$ annihilate the right vacuum while $\hat{A}^{(i)}(\overline{k})$ and $\hat{B}^{(i)}(\overline{k})$ annihilate the left vacuum. With $n^\mu x_\mu$ being equal to $-t$, in (35) we recognize the presence of non-stationary modes, which, as noted above, is characteristic of theories with non-diagonalizable Hamiltonians.

As discussed in detail in [7], the quantization procedure is also characteristic of non-diagonalizable Hamiltonians, with the commutators taking the form

$$\begin{aligned} &[A^{(i)}(\overline{k}),\hat{B}^{(j)}(\overline{k}')] = [B^{(i)}(\overline{k}),\hat{A}^{(j)}(\overline{k}')] = Z(k)\delta_{i,j}\delta^3(\overline{k}-\overline{k}'), \\ &[A^{(i)}(\overline{k}),\hat{A}^{(j)}(\overline{k}')] = 0, \qquad [B^{(i)}(\overline{k}),\hat{B}^{(j)}(\overline{k}')] = 0, \\ &[A^{(i)}(\overline{k}),B^{(j)}(\overline{k}')] = 0, \qquad [\hat{A}^{(i)}(\overline{k}),\hat{B}^{(j)}(\overline{k}')] = 0. \end{aligned} \tag{36}$$

In (36) everything except the possibly $k = |\overline{k}|$ dependent renormalization constant $Z(k)$ is fixed by kinematics (the vanishing of the $[B^{(i)}(\overline{k}),\hat{B}^{(j)}(\overline{k}')]$ commutator for instance is needed to cancel all $n^\mu x_\mu$ terms in $-4\alpha_g\langle\Omega|W_{\mu\nu}(2)|\Omega\rangle$). The constant $Z(k)$, however, will be fixed by the dynamics, with the dynamics preventing $Z(k)$ from being zero and

the above commutator algebra from being trivial. Specifically, given (36) we obtain

$$-4\alpha_g\langle\Omega|W^{\mu\nu}(2)|\Omega\rangle = \frac{2\bar{h}}{(2\pi)^3}\int_{-\infty}^{\infty} d^3k \frac{Z(k)k^\mu k^\nu}{\omega_k}, \tag{37}$$

with the factor 2 appearing in (37) since we have to sum the standard bosonic $\bar{h}\omega/2$ zero-point energy density per mode over two polarization states of two separate families of massless spin 2 modes (the $A^{(i)}(\bar{k})$ and $B^{(i)}(\bar{k})$ sectors). Then with the fermion sector generating a factor of -2 in (33) (the standard fermionic $-\bar{h}\omega$ zero-point energy density per mode as summed over negative energy states with spin up and spin down) the cancellation of the fermionic and gravitational contributions in $-4\alpha_g\langle\Omega|W^{\mu\nu}(2)|\Omega\rangle + \langle\Omega|T_{\rm M}^{\mu\nu}|\Omega\rangle = 0$ will enforce $Z(k) = 1$.

Establishing that Z is fixed by the coupling of gravity to the matter sector is our key result as it shows that gravity requires no independent quantization of its own, with its quantization strength being fixed by the consistency condition that all zero-point infinities cancel identically. To appreciate the point, it is of interest to take a more general matter source. Thus if we take the source to contain M massless gauge bosons and N massless two-component fermions (viz. $N/2$ four-component fermion modes), together they will generate $M-N$ units of $\bar{h}\omega_k$ for each $\bar{k}$. (For gauge bosons one gets $+\bar{h}\omega_k/2$ for each of two helicity states.) In this case consistency requires that Z be given by $Z = (N-M)/2$. This condition shows that Z cannot be assigned in isolation. Rather it is determined by the dynamics each time. Moreover since Z must be positive (c.f. no negative norm states) it also provides an interesting constraint on model building, namely that N must be greater than M. For the standard $SU(3)\times SU(2)\times U(1)$ model for instance, we have $M = 12$ gauge bosons and $N = 16$ two-component spinors per generation, with Z then being positive. Intriguingly, for the grand-unified gauge group $SO(10)$ one has $M = 45$ and again $N = 16$ per generation, with three generations of fermions thus being the minimum number that would make Z be positive in this case.

A second example of a dynamically determined renormalization constant may be found in two spacetime dimensions ($D = 2$). With it being the Einstein-Hilbert action that is conformally invariant in $D = 2$, to the order $\bar{h}$ of interest to us we thus couple $D = 2$ Einstein-Hilbert gravity (with $1/2\kappa_2^2$ in place of $1/16\pi G$) to a free $D = 2$ massless flat spacetime fermion. Now in $D = 2$ the classical Einstein-Hilbert action is a total divergence (the $D = 2$ Gauss-Bonnet theorem). Consequently, the associated Feynman path integral is trivial and there is no quantum scattering. However, as noted in [7, 12], the theory is not completely empty. Specifically, due to quantum ordering the quantum-mechanical Einstein-Hilbert action is a not a total divergence. (Generically, $A\partial_\mu B + B\partial_\mu A = \partial_\mu(AB) + [B,\partial_\mu A]$.) In consequence of this there are zero-point fluctuations in the gravity sector, just as needed to cancel those in the fermion sector.

The specific $D = 2$ calculation parallels the $D = 4$ case with gravitational fluctuations of the form $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ being found to satisfy a massless wave equation and with the components of $h_{\mu\nu}$ being related by the vanishing of the trace $\eta_{\mu\nu}h^{\mu\nu}$. With the momentum modes being given by $k^\mu = (\omega_k, k)$ where $\omega_k = |k|$, the general solution to the wave equation is given by [7, 12]

$$h_{00}(x,t) \;=\; \kappa_2\bar{h}^{1/2}\int \frac{dk}{(2\pi)^{1/2}(2\omega_k)^{1/2}}\left[A(k)e^{i(kx-\omega_k t)} + C(k)e^{-i(kx-\omega_k t)}\right] = h_{11}(x,t),$$

$$h_{01}(x,t) \;=\; \kappa_2\hbar^{1/2}\int\frac{dk}{(2\pi)^{1/2}(2\omega_k)^{1/2}}\left[B(k)e^{i(kx-\omega_k t)}+D(k)e^{-i(kx-\omega_k t)}\right]. \tag{38}$$

On defining

$$\begin{aligned}\langle\Omega|[C(k),B(k')]|\Omega\rangle &= -\langle\Omega|B(k)C(k)|\Omega\rangle\delta(k-k') = -f_{BC}(k)\delta(k-k'),\\ \langle\Omega|[A(k),D(k')]|\Omega\rangle &= \;\langle\Omega|A(k)D(k)|\Omega\rangle\delta(k-k') = \;\; f_{AD}(k)\delta(k-k'),\end{aligned} \tag{39}$$

where k is the spatial component of k^μ, which can be positive or negative, we find that the order $\hbar$ vanishing of $T^{\mu\nu}_{\rm GRAV}+T^{\mu\nu}_{\rm M}$ then leads to the condition

$$k[f_{BC}(k)-f_{AD}(k)] = 4\omega_k = 4|k|, \tag{40}$$

with $f_{BC}(k)-f_{AD}(k)$ being an odd function of k. As we see, the matter sector has again fixed the commutation relations for the gravitational field.

7. THE COSMOLOGICAL CONSTANT PROBLEM IN $D=2$

In dynamical generation of fermion masses one has to change the vacuum from the normal one $|N\rangle$ in which $\langle N|\bar\psi\psi|N\rangle$ is zero to a spontaneously broken one $|S\rangle$ in which $\langle S|\bar\psi\psi|S\rangle$ is non-zero. Since (20) is an operator identity it will hold in any state, and thus the cancellations required to enforce $\langle S|T^{\mu\nu}_{\rm GRAV}|S\rangle+\langle S|T^{\mu\nu}_{\rm M}|S\rangle=0$ must occur. To see how this explicitly comes about, it is instructive to consider a four-Fermi interaction in two spacetime dimensions, as that is the dimension in which the four-Fermi coupling constant g is dimensionless and the theory is conformal invariant. We introduce a flat spacetime four-Fermi action of the form $I_{\rm M}=-\int d^2x[i\hbar\bar\psi\gamma^\mu\partial_\mu\psi-(g/2)(\bar\psi\psi)^2]$, with the energy-momentum tensor $T^{\mu\nu}_{\rm M}=i\hbar\bar\psi\gamma^\mu\partial^\nu\psi-\eta^{\mu\nu}(g/2)[\bar\psi\psi]^2$ being traceless in solutions to the equation of motion, just as must be the case for conformal matter.

In the Nambu-Jona-Lasinio mean-field, Hartree-Fock approximation one looks for self-consistent states $|S\rangle$ in which $\langle S|\bar\psi\psi|S\rangle=im/g$ and $\langle S|(\bar\psi\psi-im/g)^2|S\rangle=0$. In such states the fermion equation of motion takes the form $i\hbar\gamma^\mu\partial_\mu\psi-im\psi=0$ and the mean-field energy-momentum tensor $T^{\mu\nu}_{\rm MF}$ takes the form

$$\langle S|T^{\mu\nu}_{\rm MF}|S\rangle=\langle S|i\hbar\bar\psi\gamma^\mu\partial^\nu\psi|S\rangle+\frac{m^2}{2g}\eta^{\mu\nu}, \tag{41}$$

with the mean-field approximation preserving tracelessness. With the fermion momentum modes being given by $k^\mu=(\omega_k,k)$ where $\omega_k=(k^2+m^2/\hbar^2)^{1/2}$, the quantity $\langle S|i\hbar\bar\psi\gamma^\mu\partial^\nu\psi|S\rangle$ evaluates to

$$\langle S|i\hbar\bar\psi\gamma^\mu\partial^\nu\psi|S\rangle=-\frac{\hbar}{2\pi}\int_{-\infty}^{\infty}dk\frac{k^\mu k^\nu}{\omega_k}. \tag{42}$$

In (42) we recognize mean-field energy density and pressure terms of the form

$$\rho_{\rm MF} \;=\; -\frac{\hbar}{2\pi}\left[K^2+\frac{m^2}{2\hbar^2}+\frac{m^2}{2\hbar^2}\ln\left(\frac{4\hbar^2K^2}{m^2}\right)\right],$$

$$p_{\rm MF} \;=\; -\frac{\bar{h}}{2\pi}\left[K^2+\frac{m^2}{2\bar{h}^2}-\frac{m^2}{2\bar{h}^2}\ln\left(\frac{4\bar{h}^2K^2}{m^2}\right)\right], \tag{43}$$

as conveniently cut-off at a momentum K that serves to characterize the infinities involved. In the $(m^2/2g)\eta^{\mu\nu}$ term in (41) we recognize a mean-field cosmological constant term $\Lambda_{\rm MF}=-m^2/2g$, and with $\Lambda_{\rm MF}$ evaluating to the logarithmically divergent

$$\Lambda_{\rm MF}=\frac{m^2}{4\pi\bar{h}}\ln\left(\frac{4\bar{h}^2K^2}{m^2}\right), \tag{44}$$

we obtain the gap equation $m=2\bar{h}Ke^{\pi\bar{h}/g}$. (In dynamical symmetry breaking the induced cosmological constant is infinite rather than finite – it thus appears in $(T^{\mu\nu}_{\rm M})_{\rm DIV}$ and not in $(T^{\mu\nu}_{\rm M})_{\rm FIN}$.) In terms of $\rho_{\rm MF}$, $p_{\rm MF}$ and $\Lambda_{\rm MF}$ we can write the complete mean-field $\langle S|T^{\mu\nu}_{\rm MF}|S\rangle$ as

$$\langle S|T^{\mu\nu}_{\rm MF}|S\rangle=(\rho_{\rm MF}+p_{\rm MF})U^\mu U^\nu+p_{\rm MF}\eta^{\mu\nu}-\Lambda_{\rm MF}\eta^{\mu\nu}. \tag{45}$$

Since $\langle S|T^{\mu\nu}_{\rm MF}|S\rangle$ is traceless, the various terms in (45) must obey $p_{\rm MF}-\rho_{\rm MF}-2\Lambda_{\rm MF}=0$ (in $D=2$), with all the various divergences canceling each other in the trace, just as noted in [13, 7]. Given this cancellation, we can eliminate $\Lambda_{\rm MF}$ and rewrite (45) in the manifestly traceless form

$$\begin{aligned}\langle S|T^{\mu\nu}_{\rm MF}|S\rangle &= \frac{(\rho_{\rm MF}+p_{\rm MF})}{2}\left[2U^\mu U^\nu+\eta^{\mu\nu}\right],\\ \frac{\rho_{\rm MF}+p_{\rm MF}}{2} &= \langle S|T^{00}_{\rm MF}|S\rangle=-\frac{\bar{h}}{2\pi}\left(K^2+\frac{m^2}{2\bar{h}^2}\right),\end{aligned} \tag{46}$$

with the logarithmic divergences associated with the readjustment of $\rho_{\rm MF}$ and $p_{\rm MF}$ in (43) from the massless to the massive case having completely disappeared. Finally, in order for gravity to now cancel $\langle S|T^{\mu\nu}_{\rm MF}|S\rangle$, we have to replace (40) by

$$k[f_{BC}(k)-f_{AD}(k)]=4\left[(k^2+m^2/\bar{h}^2)^{1/2}-\frac{m^2}{2\bar{h}^2(k^2+m^2/\bar{h}^2)^{1/2}}\right]. \tag{47}$$

Thus in the presence of dynamical symmetry breaking, even though the gravitational sector modes remain massless, the commutator renormalization constants in (39) readjust and become dependent on the induced fermion mass, with the renormalization constants thus being dependent on the choice of vacuum. The emergence of a behavior such as this is completely foreign to the standard canonical commutation prescription used for the matter fields, and shows how different quantization for gravity has to be. Nonetheless, with this readjustment, the vacuum cosmological constant term is completely cancelled, with conformal gravity thus being able to control the cosmological constant even after the conformal symmetry is broken.

8. THE COSMOLOGICAL CONSTANT PROBLEM IN $D = 4$

To generalize the $D = 2$ results to $D = 4$ is not immediate since in $D = 4$ the four-Fermi interaction is not conformal invariant. Rather one must work in a conformal invariant theory, which in $D = 4$ means a gauge theory. Since the renormalization procedure would introduce scaling anomaIies, to restore scale symmetry to the gauge theory one needs to be at a renormalization group fixed point. In such a case scaling would be restored but with anomalous dimensions, something first noted by Johnson, Baker and Wiley [14] in a study of quantum electrodynamics at a Gell-Mann-Low eigenvalue for the fine structure constant. In a study of dynamical symmetry breaking in this same theory it was found [15] if the dimension $d_\theta = 3 + \gamma_\theta$ of the fermion composite bilinear $\theta = \bar{\psi}\psi$ is reduced by one whole unit from its canonical value of $d = 3$ to an anomalous value of $d = 2$, the vacuum would then undergo dynamical symmetry breaking and generate a fermion mass. Specifically, with the insertion of $\bar{\psi}\psi$ into the inverse fermion propagator behaving as $\tilde{\Gamma}_\theta(p,p,0) = (-p^2/M^2)^{-1/2}$ at $\gamma_\theta = -1$, it was found that the four-Fermi value of $\varepsilon(m,4\mathrm{F}) = (i/\hbar)\int d^4p/(2\pi)^4 \mathrm{TrLn}(\gamma^\mu p_\mu - m + i\varepsilon)$, viz.

$$\varepsilon(m,4\mathrm{F}) = -\frac{\hbar}{4\pi^2}\left(K^4 + \frac{m^2K^2}{\hbar^2} - \frac{m^4}{4\hbar^4}\ln\left(\frac{4\hbar^2K^2}{m^2}\right) + \frac{m^4}{8\hbar^4}\right), \tag{48}$$

would change to

$$\begin{aligned}\varepsilon(m) &= \frac{i}{\hbar}\int\frac{d^4p}{(2\pi)^4}\mathrm{TrLn}\left[\gamma^\mu p_\mu - m\left(\frac{-p^2}{M^2}\right)^{-1/2} + i\varepsilon\right] \\ &= -\frac{\hbar K^4}{4\pi^2} + \frac{m^2M^2}{16\pi^2\hbar^3}\left[\ln\left(\frac{m^2M^2}{16\hbar^4K^4}\right) - 1\right].\end{aligned} \tag{49}$$

On setting $\varepsilon'(m) = m/g$ and $M^4 = 16\hbar^4K^4\exp(8\pi^2\hbar^3/M^2g)$, the mean-field energy density $\varepsilon(m) - m^2/2g$ thus evaluates to

$$\varepsilon(m) - \frac{m^2}{2g} = -\frac{\hbar K^4}{4\pi^2} + \frac{m^2M^2}{16\pi^2\hbar^3}\left[\ln\left(\frac{m^2}{M^2}\right) - 1\right]. \tag{50}$$

Other than the m-independent quartically divergent term (which also occurs in the gravity sector) the mean-field energy density is completely finite, with a local maximum at $m = 0$ and global minima at $m = \pm M$. With the four-Fermi $\varepsilon(m,4\mathrm{F})$ given in (48) having mass-dependent terms that are quadratically and logarithmically divergent, we see that improving the short distance-behavior of $\bar{\psi}\psi$ by one whole unit, and thus that of $\bar{\psi}\psi\bar{\psi}\psi$ by two whole units (to thereby make it non-perturbatively renormalizable) then brings the quadratic divergence down to logarithmic; with the $-m^2/2g$ term then removing the logarithmic divergence, to produce the finite terms in (50).

At the minimum, (50) takes the form

$$\varepsilon(M) - \frac{M^2}{2g} = -\frac{\hbar K^4}{4\pi^2} - \frac{M^4}{16\pi^2\hbar^3}, \tag{51}$$

in complete analog to (46). As required by (20), gravity must thus cancel the whole of (51) in $D=4$ just as it cancels the whole of (46) in $D=2$. To this end we note that the propagator $S(p)=[\gamma^\mu p_\mu - m(-p^2/M^2)^{-1/2}+i\varepsilon]^{-1}$ contained in (49) has poles, and they can be taken to be at $p^4-m^2M^2=0$ if we define the multiple-valued square root singularity appropriately. If we do the p^0 contour integration in (49) we will obtain poles at $p^2=mM$ and $p^2=-mM$. Recalling that $\langle S|\bar{\psi}\psi|S\rangle=\varepsilon'(m)$, in analog to (33) we can set

$$\varepsilon(m)-\frac{m^2}{2g}=-\frac{2\bar{h}}{(2\pi)^3}\int_{-\infty}^{\infty}d^3k \quad \Big[\quad (k^2+mM/\bar{h}^2)^{1/2}-\frac{mM}{4\bar{h}^2(k^2+mM/\bar{h}^2)^{1/2}} + \quad (k^2-mM/\bar{h}^2)^{1/2}+\frac{mM}{4\bar{h}^2(k^2-mM\bar{h}^2)^{1/2}}\Big]. \quad (52)$$

Thus at the $m=M$ minimum we obtain

$$\varepsilon(M)-\frac{M^2}{2g}=-\frac{2\bar{h}}{(2\pi)^3}\int_{-\infty}^{\infty}d^3k \quad \Big[\quad (k^2+M^2/\bar{h}^2)^{1/2}-\frac{M^2}{4\bar{h}^2(k^2+M^2/\bar{h}^2)^{1/2}} + \quad (k^2-M^2/\bar{h}^2)^{1/2}+\frac{M^2}{4\bar{h}^2(k^2-M^2/\bar{h}^2)^{1/2}}\Big]. \quad (53)$$

Comparing now with (37), we see that the cancellation of the fermionic and gravitational contributions in $-4\alpha_g\langle S|W^{\mu\nu}(2)|S\rangle+\langle S|T_{\rm M}^{\mu\nu}|S\rangle=0$ will force the gravitational sector renormalization constant $Z(k)$ in (36) to obey

$$\begin{aligned} kZ(k) &= (k^2+M^2/\bar{h}^2)^{1/2}-\frac{M^2}{4\bar{h}^2(k^2+M^2/\bar{h}^2)^{1/2}} \\ &+ (k^2-M^2/\bar{h}^2)^{1/2}+\frac{M^2}{4\bar{h}^2(k^2-M^2/\bar{h}^2)^{1/2}}, \end{aligned} \quad (54)$$

in complete analog to (47), with $Z(k)$ again being determined by the dynamics. (In (47) and (54) the numerical factor of 2 or 4 factor in the denominator is the spacetime dimension.) Additionally we note that if we were to set $M=0$ in (54) we would obtain $Z(k)=2$ rather than $Z(k)=1$, since the pole structure of the propagator $S(p)$ is that of two 4-component fermions rather than just one.

Having now obtained (53) and (54), we note that there is an alternate way to derive the structure given in (53) and (54) that is instructive in its own right. Since we are in a conformal theory we can treat the two sets of poles at $p^2=M^2$ and $p^2=-M^2$ as though they were independent degrees of freedom each with the traceless energy-tensor $T_{\mu\nu}=i\bar{\psi}\gamma_\mu\partial_\nu\psi-(1/4)\eta_{\mu\nu}\bar{\psi}\psi$ required in the broken symmetry case [1]. Recognizing $T^{00}=i\bar{\psi}\gamma^0\partial^0\psi-(1/4)\eta^{00}\bar{\psi}\psi$ to be in the generic form $T^{00}=\varepsilon(m)-(m/4)d\varepsilon(m)/dm$ for a particle of mass m, we recognize $\varepsilon(M)-(M/4)d\varepsilon(M)/d(M)+\varepsilon(iM)-(iM/4)d\varepsilon(iM)/d(iM)$ as being none other than the right-hand side of (53).

9. THE DARK MATTER PROBLEM

Since conformal gravity is a well-defined, renormalizable quantum theory, we can take matrix elements of (32) in states with an indefinite number of gravitational quanta and obtain a completely finite macroscopic limit that will be described by a classical version of (11). Classical conformal gravity has been studied by Mannheim and Kazanas [16] who found that because of the underlying conformal symmetry, the exact metric in a static, spherically symmetry geometry can be brought to the form

$$ds^2 = -B(r)dt^2 + \frac{dr^2}{B(r)} + r^2 d\Omega_2, \tag{55}$$

where the metric coefficient $B(r)$ obeys the the fourth-order equation

$$\frac{3}{B(r)}(W^0_{\ 0} - W^r_{\ r}) = \nabla^4 B = B'''' + \frac{4B'''}{r} = \frac{(rB)'''}{r} = \frac{3}{4\alpha_g B(r)}(T^0_{\ 0} - T^r_{\ r}) \equiv f(r) \tag{56}$$

without any approximation whatsoever. Exterior to a source of radius r_0 the solution to (56) is of the form

$$B(r > r_0) = 1 - \frac{2\beta}{r} + \gamma r, \tag{57}$$

with the matching of the interior and exterior solutions fixing the integration constants in (57) according to

$$2\beta = \frac{1}{6}\int_0^{r_0} dr' r'^4 f(r'), \qquad \gamma = -\frac{1}{2}\int_0^{r_0} dr' r'^2 f(r'). \tag{58}$$

Comparing with (5) and (6) we see that, as had been been noted above, we do indeed recover the Schwarzschild solution, but in addition we see that we obtain a linear potential, to thus give a departure from Newton-Einstein at large distances, i.e. in precisely the region where the dark matter problem is encountered.

Given (57), we see that a star would put out a weak gravity potential

$$V^*(r) = -\frac{\beta^* c^2}{r} + \frac{\gamma^* c^2 r}{2} \tag{59}$$

per unit solar mass. In spiral galaxies the luminous matter at a radial distance R from the galactic center is typically distributed with a surface brightness $\Sigma(R) = \Sigma_0 e^{-R/R_0}$, with the total luminosity being given by $L = 2\pi\Sigma_0 R_0^2$. If we assume that the mass distribution in a spiral galaxy is the same as that of its luminous distribution (i.e. no dark matter), then for a galactic mass to light ratio M/L, one can define the total number of solar mass units N^* in the galaxy via $(M/L)L = M = N^* M_\odot$. On integrating $V^*(r)$ over this visible matter distribution, one finds that the net centripetal acceleration due to the local luminous matter in the galaxy is given by [1]

$$\begin{aligned} \frac{v^2_{\rm LOC}}{R} &= \frac{N^*\beta^* c^2 R}{2R_0^3}\left[I_0\left(\frac{R}{2R_0}\right)K_0\left(\frac{R}{2R_0}\right) - I_1\left(\frac{R}{2R_0}\right)K_1\left(\frac{R}{2R_0}\right)\right] \\ &\quad + \frac{N^*\gamma^* c^2 R}{2R_0} I_1\left(\frac{R}{2R_0}\right)K_1\left(\frac{R}{2R_0}\right). \end{aligned} \tag{60}$$

Familiarity with Newtonian gravity would suggest that to fit galactic rotation curve data in conformal gravity one should now apply (60) as is. However there is a crucial difference between the two cases. For Newtonian gravity one uses the second-order Poisson equation $\nabla^2\phi(r) = g(r)$ and obtains a potential and force of the form

$$\phi(r) = -\frac{1}{r}\int_0^r dr' r'^2 g(r') - \int_r^\infty dr' r' g(r'), \qquad \frac{d\phi(r)}{dr} = \frac{1}{r^2}\int_0^r dr' r'^2 g(r'). \tag{61}$$

As such, the import of (61) is that even though $g(r)$ could continue globally all the way to infinity, the force at any radial point r is determined solely by the material in the local $0 < r' < r$ region. In this sense Newtonian gravity is local in character, since to explain a gravitational effect in some local region one only needs to consider the material in that region. Thus in Newtonian gravity, if one wishes to explain the behavior of galactic rotation curves through the use of dark matter, one must locate the dark matter where the problem is and not elsewhere, i.e. within the galaxies themselves.

However, this familiar property of Newtonian gravity is not generic to any theory of gravity. In particular if we define $h(r) = c^2 f(r)/2$, the conformal gravity potential associated with (56) will obey the fourth-order Poisson equation $\nabla^4\phi(r) = h(r)$, with general solution

$$\begin{aligned}\phi(r) &= -\frac{r}{2}\int_0^r dr' r'^2 h(r') - \frac{1}{6r}\int_0^r dr' r'^4 h(r') - \frac{1}{2}\int_r^\infty dr' r'^3 h(r') - \frac{r^2}{6}\int_r^\infty dr' r' h(r') \\ \frac{d\phi(r)}{dr} &= -\frac{1}{2}\int_0^r dr' r'^2 h(r') + \frac{1}{6r^2}\int_0^r dr' r'^4 h(r') - \frac{r}{3}\int_r^\infty dr' r' h(r').\end{aligned} \tag{62}$$

As we see, this time we do find a global contribution to the force coming from material in the $r < r' < \infty$ region that is beyond the radial point of interest. Hence in conformal gravity one cannot ignore the rest of the universe, with a test particle in orbit in a galaxy being able to sample both the local field due to the matter in the galaxy and the global field due to the rest of the matter in the Universe. Unlike Newtonian gravity then, conformal gravity is an intrinsically global theory.

The contribution that the rest of the Universe provides consists of two components, the homogeneous cosmological background and the inhomogeneities in it. The homogeneous background can be described by a Roberston-Walker (RW) geometry, while large scale inhomogeneities are typically in the form of large gravitationally bound systems such as clusters and superclusters. Since the RW metric is conformal to flat, and since the Weyl tensor vanishes identically in a such a geometry, the cosmological background is characterized by a geometry in which $W^{\mu\nu}$ of (11) (and thus the cosmological $T_{\rm M}^{\mu\nu}$) vanish identically. However, since localized inhomogeneities have a non-vanishing Weyl tensor, the inhomogeneities contribute to the integrals in (62) that extend out to infinity beyond the galaxy of interest. The inhomogeneities contribute to the particular integral solution to (11) given in (62) (both $\nabla^4 B(r)$ and $f(r)$ non-zero), while the homogeneous background contributes to the complementary function (both $\nabla^4 B(r)$ and $f(r)$ zero).

In order for the background cosmology to contribute non-trivially, we note that even though we need the background $T_{\rm M}^{\mu\nu}$ to vanish (since the RW $W^{\mu\nu}$ vanishes), we would need $T_{\rm M}^{\mu\nu}$ to vanish non-trivially if it is to have any content. As shown in [1], such a non-trivial vanishing can be achieved by an interplay between the positive contribution of the

matter sources (c.f. $\Sigma\hbar\omega a^\dagger a$) and the negative contribution of the gravitational field that occurs if the 3-curvature K of the Universe is negative, with gravity providing negative energy density. In [1] it was shown that with such a cosmology one could then fit the accelerating universe Hubble plot data without the need for any fine-tuning of parameters or for any of the cosmological dark matter required in the standard theory. (Unlike the standard $\Omega_{\rm K}=0$ cosmology, which is fine-tuned to only accelerate at late redshift, with its negative K the conformal cosmology naturally accelerates at all redshifts.)

Since cosmology is written in comoving Hubble flow coordinates while rotation curves are measured in galactic rest frames, to ascertain the impact of cosmology on rotation curves one needs to transform the RW metric to static coordinates. As noted in [16], the transformation

$$\rho=\frac{4r}{2(1+\gamma_0 r)^{1/2}+2+\gamma_0 r},\qquad \tau=\int dtR(t) \tag{63}$$

effects the metric transformation

$$\begin{aligned}&-(1+\gamma_0 r)c^2dt^2+\frac{dr^2}{(1+\gamma_0 r)}+r^2d\Omega_2=\\&\frac{1}{R^2(\tau)}\left(\frac{1+\gamma_0\rho/4}{1-\gamma_0\rho/4}\right)^2\left[-c^2d\tau^2+\frac{R^2(\tau)}{[1-\gamma_0^2\rho^2/16]^2}\left(d\rho^2+\rho^2d\Omega_2\right)\right].\end{aligned} \tag{64}$$

Recognizing (64) to be conformal to a topologically open RW metric with 3-curvature $K=-\gamma_0^2/4$, and recalling that in metrics conformal to RW the tensor $W^{\mu\nu}$ still vanishes, we see that in the rest frame of a galaxy the negative K global cosmology found in [1] acts like a universal linear potential with cosmological strength $\gamma_0/2=(-K)^{1/2}$.

In the weak gravity limit one can add this global potential on to (60), with the total centripetal acceleration then being given by [17]

$$\frac{v_{\rm TOT}^2}{R}=\frac{v_{\rm LOC}^2}{R}+\frac{\gamma_0 c^2}{2}. \tag{65}$$

In [17] (65) was used to fit the galactic rotation velocities of a sample of 11 spiral galaxies, and good fits were found, with the two universal linear potential parameters being fixed to the values

$$\gamma^*=5.42\times10^{-41}{\rm cm}^{-1},\qquad \gamma_0=3.06\times10^{-30}{\rm cm}^{-1}. \tag{66}$$

The value obtained for γ^* entails that the linear potential of the Sun is so small that there are no modifications to standard solar system phenomenology, with the values obtained for $N^*\gamma^*$ and γ_0 being such that one has to go to galactic scales before their effects can become as big as the Newtonian contribution.

However, as we had noted above, there is a contribution due to inhomogeneities in the cosmic background that we need to include too. These inhomogeneities would typically be clusters and superclusters and would be associated with distance scales between 1 Mpc and 100 Mpc or so. Without knowing anything other than that about them, we see from (62) that for calculating potentials at galactic distance scales (viz. scales much less

than cluster scales) the inhomogeneities would contribute constant and quadratic terms multiplied by integrals that are evaluated between end points that do not depend on the galaxy of interest, to thus be constants. Thus we augment (65) to

$$\frac{v_{\rm TOT}^2}{R} = \frac{v_{\rm LOC}^2}{R} + \frac{\gamma_0 c^2}{2} - \kappa c^2 R, \tag{67}$$

with asymptotic limit

$$\frac{v_{\rm TOT}^2}{R} \to \frac{N^* \beta^* c^2}{R^2} + \frac{N^* \gamma^* c^2}{2} + \frac{\gamma_0 c^2}{2} - \kappa c^2 R. \tag{68}$$

Armed with (67) Mannheim and O'Brien [18, 19] set out to update the earlier fits of [17] and apply the conformal theory to a sample of 110 galaxies that had become available in the interim (a varied and broad sample that includes both high and low surface brightness galaxies and dwarfs). In making such fits the only parameter that can vary from one galaxy to the next is the galactic disk mass to light ratio as embodied in N^*, with the parameters γ^*, γ_0 and κ needing to be universal and not have any dependence on a given galaxy at all. To model the contribution of the luminous matter known photometric surface brightness data parameters were used. The fits are thus highly constrained, one parameter per galaxy fits with all photometric input data being known, with everything else being universal, and with no dark matter being assumed.

Now since the κ-dependent term had not been used in the fits given in [17], one would immediately expect that it would be too small to be significant. However, because the 110 galaxy sample is so big, it contains galaxies whose rotation velocity data go out to radial distances much larger than the ones that had previously been considered. These data are thus sensitive to the distance-dependent $-\kappa c^2 R$ term present in (67), with the fitting underscoring the value of working with a large data sample. The fitting to the complete 110 galaxy sample is reported in [19], with the fitting to the 20 largest galaxies (viz. those that are most sensitive to the $-\kappa c^2 R$ term) being reported here and in [18]. The fitting shows that without any galactic dark matter (67) captures the essence of the data for the entire 110 galaxy sample, with the parameters γ^* and γ_0 continuing to take the values given in (66), and with κ being found to take the value

$$\kappa = 9.54 \times 10^{-54}\ {\rm cm}^{-2}. \tag{69}$$

In the figures we present the actual fitting to the 20 galaxy sample with all details being given in [18, 19]. In the figures the rotational velocities and errors (in km sec^{-1}) are plotted as a function of radial distance (in kpc). For each galaxy we exhibit the contribution due to the luminous Newtonian term alone (dashed curve), the contribution from the two linear terms alone (dot dashed curve), the contribution from the two linear terms and the quadratic terms combined (dotted curve), with the full curve showing the total contribution. Because the data go out to such large distances the data are sensitive to the rise in velocity associated with the linear potential terms, and it is here that the quadratic term acts to actually arrest the rise altogether (dotted curve) and cause all rotation velocities to ultimately fall. Moreover, since v^2 cannot be negative, beyond a distance R of order $\gamma_0/\kappa = 3.21 \times 10^{23}$ cm or so there could no longer be any bound

galactic orbits, with galaxies thus having a natural way of terminating, and with global physics thus imposing a natural limit on the size of galaxies. To illustrate this we plot the rotation velocity curve for UGC 128 over an extended range.

It is important to appreciate that the fits provided by conformal gravity (and likewise those provided by other alternate theories such as Milgrom's MOND theory and Moffat's MSTG theory [20]) are predictions. Specifically, for all these theories the only input one needs is the photometric data, and the only free parameter is the M/L ratio for each given galaxy, with rotation velocities then being determined. That these highly constrained alternate theories all work is because not only do they each possess an either derived or postulated underlying universal scale (a derived $\gamma_0 = 2(-K)^{1/2} = 3.06 \times 10^{-30}$ cm^{-1} for conformal gravity, $a_0/c^2 = 1.33 \times 10^{-29}$ cm^{-1} for MOND and $G_0 M_0/r_0^2 c^2 = 7.67 \times 10^{-29}$ cm^{-1} for MSTG), all of the 110 galaxies in the sample possess it too. Specifically, despite the huge variation in luminosity and surface brightness across the 110 galaxy sample, within one order of magnitude the measured values of the centripetal accelerations $(v^2/c^2R)_{\rm last}$ at the last data point in each galaxy are all found to cluster around a value of 3×10^{-30} cm^{-1} or so. For the 20 large galaxy sample for instance the values for v^2/c^2R all lie within the narrow range $(0.97 - 5.83) \times 10^{-30}$ cm^{-1}.

It should also be noted that while the fits provided by conformal gravity are predictions, in contrast, dark matter fitting to galactic data works quite differently. There one first needs to know the velocities so that one can then ascertain the needed amount of dark matter, i.e. in its current formulation dark matter is only a parametrization of the velocity discrepancies that are observed and is not a prediction of them. Dark matter theory has yet to develop to the point where it is able to predict rotation velocities given a knowledge of the luminous distribution alone (or explain the near universality found for $(v^2/c^2R)_{\rm last}$). Thus dark matter theories, and in particular those theories that produce dark matter halos in the early universe, are currently unable to make an a priori determination as to which halo is to go with which particular luminous matter distribution, and need to fine-tune halo parameters to luminous parameters galaxy by galaxy. In the FNW CDM simulations [21] for instance, one finds generic spherical halo profiles close in form to $\sigma(r) = \sigma_0/[r(r+r_0)^2]$ (as then cut off at cr_0), but with the halo parameters σ_0, r_0 and c needing to be fixed galaxy by galaxy. In addition to the galactic mass to light ratios, this requires 330 further parameters for the 110 galaxy sample (or a further 220 parameters for isothermal halo type models). No such fine-tuning shortcomings appear in conformal gravity, and if standard gravity is to be the correct description of gravity, then a universal formula akin to the one given in (67) and the existence of the universal γ_0 and κ parameters would need to be derived by dark matter theory.

The conformal gravity fits are also noteworthy in that conformal gravity was not at all developed for the purpose of addressing the dark matter problem. Rather, it was first advocated by the present author [22] solely because it has a symmetry that could address the cosmological constant problem. However, once the starting action of (9) is assumed, one can then proceed purely deductively and derive the rotation curve formula given in (67), a thus purely theoretical first principles approach. Moreover, since our study of (67) then establishes that global physics has an influence on local galactic motions, the invoking of dark matter in galaxies could potentially be nothing more than an attempt to describe global physics effects in purely local galactic terms.

10. CONNECTING CONFORMAL AND EINSTEIN GRAVITY

In a recent paper 't Hooft [2] has found an interesting connection between Einstein gravity and conformal gravity. In standard treatments of quantum Einstein gravity one makes a perturbative expansion in the metric and generates multi-loop Feynman diagrams. Each perturbative order requires a new counter-term, with the nth-order one being a function of the nth-power of the Riemann tensor and its contractions. With the series not terminating, Einstein gravity is rendered non-renormalizable.

In his paper 't Hooft proposes a very different approach, one that is highly nonlinear. Specifically, instead of evaluating the path integral as a perturbative series in the metric components $g_{\mu\nu}(x)$, he instead proposes to treat the conformal factor in the metric as an independent degree of freedom. Specifically, he makes a conformal transformation on the (non-conformal invariant) Einstein-Hilbert action of the form $g_{\mu\nu}(x) = \omega^2(x)\hat{g}_{\mu\nu}(x)$, to obtain

$$I_{\rm EH} = -\frac{1}{16\pi G}\int d^4x(-\hat{g})^{1/2}\left(\omega^2\hat{R}^{\alpha}{}_{\alpha} - 6\hat{g}^{\mu\nu}\partial_\mu\omega\partial_\nu\omega\right), \tag{70}$$

with everything in $I_{\rm EH}$ now being evaluated in a geometry with metric $\hat{g}_{\mu\nu}(x)$ [23]. Then instead of taking the path integral measure to be of the standard $Dg_{\mu\nu}$ form, 't Hooft proposes that it be taken to be of the form $D\omega D(g_{\mu\nu}/\omega^2) = D\omega D\hat{g}_{\mu\nu}$ instead.

The utility of this approach is that since the ω dependence in $I_{\rm EH}$ is quadratic, the $D\omega$ path integral can be done analytically. However, in order for the path integral to be bounded one needs to give ω an imaginary part. With this choice, the ω path integral will generate an effective action $I_{\rm EFF}$ of the form Tr $\ln[\hat{g}^{\mu\nu}\hat{\nabla}_\mu\hat{\nabla}_\nu + (1/6)\hat{R}^{\alpha}{}_{\alpha}]$, and after dimensional regularization is found not to generate an infinite set of divergent terms at all, but rather to only generate just one divergent term, viz. the logarithmically divergent

$$I_{\rm EFF} = \frac{C}{120}\int d^4x(-\hat{g})^{1/2}[\hat{R}^{\mu\nu}\hat{R}_{\mu\nu} - \frac{1}{3}(\hat{R}^{\alpha}{}_{\alpha})^2], \tag{71}$$

with C being the very same logarithmically divergent constant that had appeared in (16).

In (71) we immediately recognize the conformal gravity action. Since the action in (70) is the same action as that obeyed by a conformally coupled scalar field, and the ω path integral measure is the same as that of a scalar field, everything is conformal, and the ω path integral can only generate a conformal invariant effective action – hence (71). Through the unusual treatment of the conformal factor we thus find a connection between Einstein gravity and conformal gravity.

From the perspective of Einstein gravity, the utility of (71) is that since conformal gravity is renormalizable, the subsequent $D\hat{g}_{\mu\nu}$ integration of $I_{\rm EFF}$ should not generate any additional counter-terms, while the non-leading terms contained in Tr $\ln[\hat{g}^{\mu\nu}\hat{\nabla}_\mu\hat{\nabla}_\nu + (1/6)\hat{R}^{\alpha}{}_{\alpha}]$ would generate contributions to the path integration that could potentially be finite. One still has to deal with the divergent C term in (71), and rather than have it renormalized (say by adding on an intrinsic conformal $I_{\rm W}$ term), 't Hooft explores the possibility that C remain uncanceled, and we refer the reader to his paper for details. (With C appearing with the same overall sign in (16) and (71), and with a gauge field path integration over its kinetic energy having the same sign too [2], an interplay between fermionic and bosonic fields cannot cancel C – with massless

superpartner fields not being able to effect the same cancellation in a curved background that they can achieve in a flat one.)

While we thus generate a conformal action if we start with an Einstein action, from the perspective of pure conformal gravity, conformal invariance prevents one from including an Einstein term in the fundamental action at all. However, in a pure conformal theory one still needs to use a measure that is conformal invariant. Now the metric itself is not conformal invariant, and hence neither is the $Dg_{\mu\nu}$ path integration measure. However, the quantity $g_{\mu\nu}/(-g)^{1/4}$ is conformal invariant and thus so is an integration measure of the form $D(g_{\mu\nu}/(-g)^{1/4})$ (and analogously $D((-g)^{3/16}\bar{\psi})D((-g)^{3/16}\psi)$ for fermions). In order to simplify the measure one would like to work in a conformal gauge in which the determinant of the metric is fixed to a convenient value such as one. And while it needs to be explored in detail, it is possible that adding an Einstein term to the conformal action and taking the measure to be of the form $D(-g)^{1/8}D(g_{\mu\nu}/(-g)^{1/4})$ might then serve as an appropriate conformal gauge fixing procedure. The fact that there would only be nine independent $D(g_{\mu\nu}/(-g)^{1/4})$ terms parallels the fact that the perturbative $I_{\rm W}$ of (22) only depends on the traceless 9-component $K^{\mu\nu} = h^{\mu\nu} - (1/4)\eta^{\mu\nu}\eta_{\alpha\beta}h^{\alpha\beta}$. Additionally, the fact that one needs to give ω an imaginary part in order to obtain a well-defined path integral parallels the structure we found for the conformal theory, with unitarity being achieved by having the gravitational field be anti-Hermitian.

11. SUMMARY

In this review we have presented the case that can be made for conformal gravity. In particular we have shown how it can quite naturally handle some of the most troublesome problems in physics, the quantum gravity problem, the vacuum energy problem, and the dark matter problem, being able to do so in the four spacetime dimensions for which there is evidence. As detailed in [1] much more still needs to be done: anisotropy of the cosmic microwave background, large scale structure, cluster dynamics and lensing by clusters (especially in light of the global $-\kappa c^2 R$ term in (67)), orbit decays of binary pulsars and gravity waves, solving the primordial deuterium problem that conformal nucleosynthesis has. For all of these applications we only need to consider the particle contribution to the finite (32), with the contribution of the vacuum sector including the cosmological constant having been taken care of by (31). All of these issues should eventually prove definitive for the conformal theory, especially since it has none of the freedom associated with the difficult to pin down and still highly elusive dark matter and dark energy present in the standard theory. The highly constrained conformal gravity fits to galactic rotation curves have as yet no parallel in dark matter theory where parameters need to be fine-tuned galaxy by galaxy, and its natural solution to the cosmological constant problem has as yet no parallel in standard cosmology where Λ needs to be fine-tuned to an unbelievable degree. To conclude, we note that at the beginning of the 20th century studies of black-body radiation on microscopic scales led to a paradigm shift in physics. Thus it could that at the beginning of the 21st century studies of black-body radiation, this time on macroscopic cosmological scales, might be presaging a paradigm shift all over again.

REFERENCES

1. P. D. Mannheim, *Prog. Part. Nucl. Phys.* **56**, 340 (2006).
2. G. 't Hooft, *Probing the small distance structure of canonical quantum gravity using the conformal group*, arXiv:1009.0669v2 [gr-qc]
3. C. M. Bender and P. D. Mannheim, *Phys. Rev. Lett.* **100**, 110402 (2008).
4. C. M. Bender and P. D. Mannheim, *Phys. Rev. D* **78**, 025022 (2008).
5. T. Eguchi and H. Sugawara, *Phys. Rev. D* **10**, 4257 (1974).
6. P. D. Mannheim, *Phys. Rev. D* **14**, 2072 (1976).
7. P. D. Mannheim, *Comprehensive solution to the cosmological constant, zero-point energy, and quantum gravity problems*, arXiv:0909.0212v5 [hep-th], General Relativity and Gravitation, in press.
8. C. M. Bender, *Rep. Prog. Phys.* **70**, 947 (2007).
9. C. M. Bender, M. V. Berry and A. Mandilara, *J. Phys. A: Math. Gen.* **35**, L467 (2002).
10. C. M. Bender and P. D. Mannheim, *Phys. Lett. A* **374**, 1616 (2010).
11. P. D. Mannheim, *PT symmetry as a necessary and sufficient condition for unitary time evolution*, arXiv:0912.2635v1 [hep-th]. A. Mostafazadeh, *Pseudo-Hermitian representation of quantum mechanics*, arXiv:0810.5643v3 [quant-ph].
12. P. D. Mannheim, *Intrinsically quantum-mechanical gravity and the cosmological constant problem*, arXiv:1005.5108v3 [hep-th].
13. P. D. Mannheim, *Dynamical symmetry breaking and the cosmological constant problem*, Proceedings of the 34th International Conference in High Energy Physics (ICHEP08), Philadelphia, 2008, eConf C080730. (arXiv:0809.1200 [hep-th])
14. K. Johnson, M. Baker and R. Willey, *Phys. Rev.* **136**, B1111 (1964). K. Johnson, R. Willey and M. Baker, *Phys. Rev.* **163**, 1699 (1967). M. Baker and K. Johnson, *Phys. Rev.* **183**, 1292 (1969); *Phys. Rev. D* **3**, 2516 (1971); *Phys. Rev. D* **3**, 2541 (1971). K. Johnson and M. Baker, *Phys. Rev. D* **8**, 1110 (1973).
15. P. D. Mannheim, *Phys. Rev. D* **10**, 3311 (1974); *Phys. Rev. D* **12**, 1772 (1975); *Nucl. Phys. B* **143**, 285 (1978).
16. P. D. Mannheim and D. Kazanas, *Ap. J.* **342**, 635 (1989); *Gen. Rel. Gravit.* **26**, 337 (1994).
17. P. D. Mannheim, *Ap. J.* **479**, 659 (1997).
18. P. D. Mannheim and J. G. O'Brien, *Impact of a global quadratic potential on galactic rotation curves*, arXiv:1007.0970v2 [astro-ph.CO]
19. P. D. Mannheim and J. G. O'Brien, *Fitting galactic rotation curves with conformal gravity and a global quadratic potential*, arXiv:1011.3495v2 [astro-ph.CO]
20. M. Milgrom, *Ap. J.* **270**, 365, 371, 384 (1983). J. R. Brownstein and J. W. Moffat, *Ap. J.* **636**, 721 (2006).
21. J. F. Navarro, C. S. Frenk and S. D. M. White, *Ap. J.* **462**, 563 (1996); *Ap. J.* **490**, 493 (1997).
22. P. D. Mannheim, *Gen. Rel. Gravit.* **22**, 289 (1990).
23. In deriving (70) we note that if set $g_{\mu\nu}(x) = \omega^2(x)\hat{g}_{\mu\nu}(x)$ we obtain $(-g)^{1/2}R^{\alpha}{}_{\alpha}(g_{\mu\nu}) = (-\hat{g})^{1/2}[\omega^2\hat{R}^{\alpha}{}_{\alpha}(\hat{g}_{\mu\nu}) + 6\bar{g}^{\alpha\beta}\omega\hat{\nabla}_{\alpha}\hat{\nabla}_{\beta}\omega]$ where the $\hat{\nabla}_{\alpha}$ derivatives are evaluated in a geometry with metric $\hat{g}_{\mu\nu}(x)$. An integration by parts then yields (70).

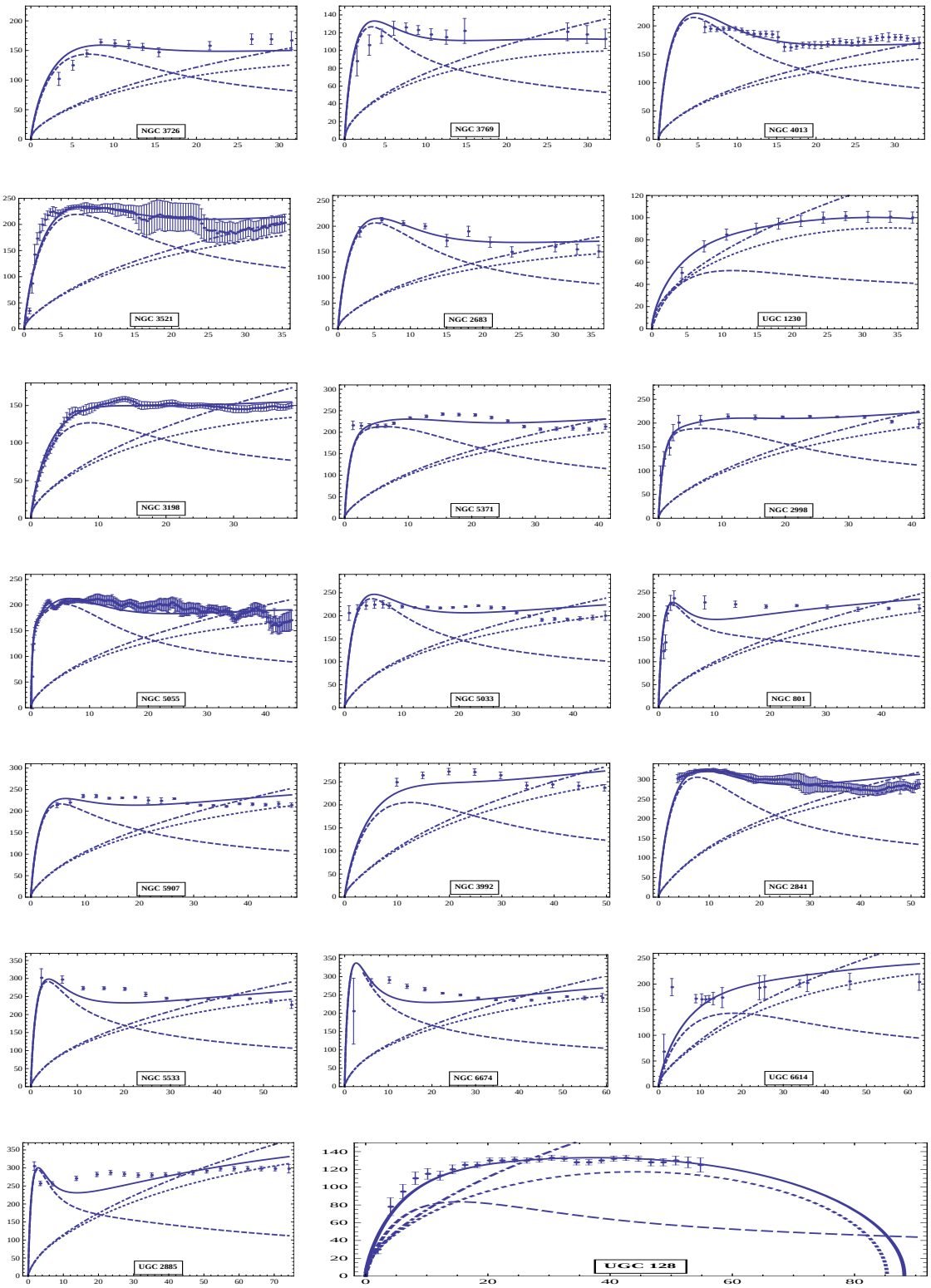


FIGURE 1. Fitting to the rotational velocities in km sec^{-1} as a function of radial distance in kpc

Problems with Conformal Gravity

Hans C. Ohanian
Department of Physics, University of Vermont, Burlington, VT 05405-0125, USA

Mannheim's theory of conformal gravity (see these Proceedings) is a geometric theory, which relies on the usual assumption that ideal particles, of negligible size and mass, move along the geodesics of a curved spacetime geometry generated by nearby mass distributions. However, in such geometric theories, the assumed geodesic motion of ideal particles does not necessarily imply the geodesic motion of real, extended bodies. Deviations from geodesic motion can arise from the finite size of a body, which makes the body sensitive to gradients in the gravitational fields (tidal forces). And deviations can also arise from the finite mass of the body, which gives it a gravitational self-energy, which can alter the ratio of gravitational mass to inertial mass and the rate of free fall (this does not happen in general relativity, but it does happen in all other theories of gravitation I know of; for instance, deviation from geodesic motion because of gravitational self-energy is a well-known feature of the Brans-Dicke theory). Conformal gravity is especially susceptible to deviations from geodesic motion arising from the finite sizes of bodies because of the mechanism by which it generates the $1/r$ gravitational potential. With the usual correspondence between metric tensor and nonrelativistic potential, $-g_{00} = B(r) \simeq 1 + 2\Phi(r)$, the field equation in conformal gravity for a spherically symmetric gravitational field is the fourth-order Poisson equation

$$\nabla^4 \Phi(r) = \frac{3}{8\alpha_g} \frac{(T_0^0 - T_r^r)}{B(r)} \tag{1}$$

For an ideal pointlike source with $(T_0^0 - T_r^r)\ \alpha\ \delta^3(\vec{x})$, the exterior solution of this equation is

$$\Phi(r)\ \alpha\ r \tag{2}$$

Thus, the "natural" potential for a theory with this fourth-order Poisson equation is not a $1/r$ potential, but a linear r potential (that is, not an inverse-square force, but a constant radial force). However, for a macroscopic spherical source of finite radius R—such as a bowling ball or a planet—the exterior solution of the fourth-order Poisson equation is [1]

$$\Phi(r) = -\frac{3}{4\alpha_g}\frac{r}{16\pi}\int_0^r \frac{(T_0^0 - T_r^r)}{B(r')}dV' - \frac{1}{4\alpha_g}\frac{1}{16\pi r}\int_0^r \frac{(T_0^0 - T_r^r)}{B(r')}r'^2 dV' \tag{3}$$

This potential includes a linear r term, but also a $1/r$ term, which Mannheim proposes to identify with the familiar $1/r$ Newtonian gravitational potential. In Eq. (3) this term $1/r$ is really a kind of quasi-multipole term that arises when a point source with a linear r potential [see Eq. (2)] is replaced by a sphere. In contrast to the multipoles in Newtonian potential theory (e.g., quadrupoles), the quasi-multipole in Eq. (3) has no angular dependence [2].

If the source function $(T_0^0 - T_r^r)$ is reasonably smooth, then the second term on the right side of Eq. (3) is smaller than the first term by a factor of about R^2/r^2 , which makes it negligible at a distance $r >> R$ from the source. Mannheim seeks to evade this conclusion by hypothesizing that the source function $(T_0^0 - T_r^r)$ has violent microscopic singularities, or "discrete sources",

at unspecified locations within atoms or nuclei, and that at these singularities the integrand $(T_0^0 - T_r^r)r^2/B(r)$ of the second term in Eq. (3) picks up contributions much larger than the integrand $(T_0^0 - T_r^r)/B(r)$ of the first term [3]. The $1/r$ potential then becomes much larger than the linear r potential, in agreement with observations in our neighborhood; and the linear r potential shows up only at large distances, in the outer reaches of galaxies, where it serves to solve the dark-mass problem.

I have several objections to this roundabout way of generating a $1/r$ potential. My two main objections are that *i*) the predicted motion of real test bodies deviates from the geodesics of the metric tensor $g_{\mu\nu}$ a deviation that can be easily detected by comparing planetary motion with the motion of light signals, and *ii*) the predicted accelerations of free fall of test bodies are not equal, and the differences are much larger than the experimental limits set by the Eötvös experiments. The following arguments explain these objections:

i) When a spherical mass m of finite size moves in the gravitational field of a spherical mass M of finite size,there is a discrepancy between the actual potential energy of these two masses and the potential energy calculated from the gravitational potential $\Phi(r)$ given by Eq. (3). For an ideal pointlike particle of mass m, the potential energy is of course simply $m\Phi(r)$ that is, Eq. (3) gives the correct potential energy. However, for a spherical test body, with a mass distribution of finite size, the total potential energy includes not only the $1/r$ term generated by the mass distribution M, but also an extra $1/r$ term generated by the mass distribution m .The $1/r$ terms in the total potential energy then become

$$U(r) = "(r)terms" - \frac{1}{4\alpha_g}\frac{m}{16\pi r}\int_M \frac{(T_0^0 - T_r^r)}{B(r')}r'^2 dV' - \frac{1}{4\alpha_g}\frac{M}{16\pi r}\int_m \frac{(T_0^0 - T_r^r)}{B(r')}r'^2 dV' \qquad (4)$$

where the first integral on the right side is over the mass distribution m and the second over the mass distribution. Under Mannheim's hypothesis, the integrand of each of the two integrals on the right side of Eq. (5) is a sum over discrete microscopic singularities, and the first integral gives a result proportional to M whereas the second gives a result proportional to m. Accordingly, the two $1/r$ term in Eq. (4) are then equal, and the effective $1/r$ potential energy is twice as large as expected from naïve application of Eq. (3).

This means that although conformal gravity has a geometrical basis in that the motion of ideal pointlike particles proceeds along the geodesics of the curved spacetime geometry, the motion of test bodies of finite size includes a significant correction that makes the bodies deviate from these geodesics (this deviation is somewhat analogous to the deviation from Keplerian motion we encounter in Newtonian dynamics if our test body is endowed with a quadrupole moment). Although the underlying theory is geometric, the effective, "emergent," theory is nongeometric.

This deviation from geodesic motion is readily detectable by comparing the motion of test bodies with the motion of light signals. The extra term in Eq. (4) does not exist for light signals, because light waves do not contain singularities a la Mannheim, and a wavepacket does not generate a $1/r$ potential. Hence the doubling of the $1/r$ term in Eq. (4) is absent, and light signals simply follow the geodesics of the curved spacetime geometry with $-g_{00} = B(r) \simeq 1 + 2\Phi(r)$, $g_{rr} = 1/B(r)$. The coupling of a light signal to the gravitational field is therefore effectively half as large as that of a test body (such as a planet), and the deflection of light by the Sun is predicted to be 0.85 arcseconds, instead of the observed 1.7 arcseconds. Likewise, the Shapiro time retardation for light is predicted to be half its standard value. Both the light deflection and the time retardation have been measured and confirmed with high precision ± 2, parts in 10^4 for the former and parts in for the latter. This is clear observational evidence against conformal gravity.

ii) My second argument relies on the Eötvös experiments. Given the evidence provided by the above comparison of the motion of test bodies vs. light signals, any additional argument

may seem gratuitous. However, this second argument highlights some important general points that apply not only to conformal gravity, but also to other proposed theories of gravity, whose inventors might benefit from a better understanding of the full implications of the Eötvös experiments.

Although the Eötvös experiment is usually viewed as a test of geodesic motion (that is, equal accelerations of free fall for all bodies, or what is called the weak equivalence principle), the experiment also provides information about the source of the gravitational interaction. Traditionally, the experimental results are described by expressing the equation of motion in a nonrelativistic form, valid for small deviations from flat spacetime and low speeds,

$$m_I \frac{d^2x^k}{dt^2} = -m_G \frac{\partial \Phi}{\partial x^k} \tag{5}$$

This corresponds to geodesic motion if the inertial and gravitational masses are equal $m_I = m_G$, Note that here the gravitational mass m_G is the "passive" mass (the receptor of gravitational force). Logically, this is distinct from the "active" mass m'_G (the source of gravitational force). However, in any theory that includes momentum conservation (that is, in all Lagrangian field theories, such as conformal gravity), the active mass necessarily equals the passive mass. This is immediately obvious from the equality of action and reaction forces between two masses m and M, which requires $m_G M'_G = m'_G M_G$ and therefore implies that the ratio $\frac{m_G}{m'_G}$ is a universal constant, which we can set equal to 1. Accordingly, measurements of the ratio $\frac{m_G}{m_I}$ obtained from Eötvös experiments are equally valid for the ratio $\frac{m'_G}{m'_I}$. The Eotvos experiments therefore determine the gravitational sourcestrengths of mass samples, even though the experiments do not directly measure the gravitational force exerted by these mass samples. This has crucial consequences for conformal gravity, because the source strengths assumed by conformal gravity contradict the results of the Eötvös experiments. The available data from a wide range of samples of materials tested in the experiments establish that to a precision of 3 parts in 1013 the gravitational mass—or the source strength of the $1/r$ potential—is equal to the inertial mass. For conformal gravity, this means that the coefficient of the $1/r$ potential in Eq. (3) should be equal to Gm_I that is, it should be equal to $G\int(-T_0^0)dV'$:

$$\frac{1}{4\alpha_G}\frac{1}{16\pi}\int_0^R \frac{(T_0^0 - T_r^r)}{B(r')} r'^2 dV' = G\int_0^R (-T_0^0)dV' \tag{6}$$

For a detailed examination of this relationship, we can separate the energy momentum tensor into several contributions corresponding to the several constituents of the atom and its nucleus: electrons, protons, neutrons, strong nuclear binding energy, atomic and nuclear electrostatic energy, etc. The ratio of the separate gravitational and inertial mass contributions for each separate atomic and nuclear constituent is known from the analysis of the available Eötvös data. Thus, Will [4] lists the results $|\Delta m_G - \Delta m_I| / \Delta m_I < 10^{-12}, 5\times10^{-10}, 4\times10^{-10}, 6\times10^{-10}$ respectively, for the upper limits on gravitational and inertial mass differences associated with baryons, strong nuclear binding energy (gluon field energy), nuclear electrostatic energy, and nuclear magnetostatic energy. To achieve agreement with these experimental results, the equality between the integrals on the right and left sides of Eq. (6) must be satisfied to within these experimental limits for each separate atomic and nuclear constituent. For the electron, proton, and neutron constituents, the equality is simply a hypothesis about the fixed interior structure of these particles, which might be regarded as terra incognita about which Mannheim can hypothesize at will (although it is a deplorable deficiency of conformal gravity that the equality of gravitational and inertial masses must be inserted "by hand" for each of these particles). But for electrostatic, magnetostatic, and nuclear binding energies, the integrals in Eq. (6) depend on the individual

nucleus and its quantum-mechanical wavefunctions and the energy-momentum tensors of electromagnetic and gluon fields. The value of each integral is calculable from the known nuclear and nucleon structure, and it cannot be stipulated by hypothesis.

For instance, the electrostatic contribution can be calculated fairly simply, by interpreting the square of the proton wavefunction as a charge density. For a rough calculation, it actually suffices to use the liquid-drop model, with a uniform charge density, which generates a radial electric field (alternatively, it is possible to use the charge density determined by scattering experiments; this confirms that the nuclear charge density is nearly uniform). For a radial electric field $T^r_r = -T^0_0$ and since the energy density $-T^0_0$ in the electric field is strictly positive, the integrands on the left and right sides of Eq. (6) have opposite signs, and they cannot be equal. That is, in conformal gravity, nuclear electrostatic energy contributes negative gravitational mass and produces a deviation $|\Delta m_G - \Delta m_I| / \Delta m_I > 1$ between the gravitational and inertial mass contributions attributable to the electrostatic energy of the nucleus. This deviation exceeds the experimental limit listed above by many orders of magnitude [5]. Similar calculations, with similar results, can be performed for nuclear magnetostatic energy and probably also for gluon field energy.

I believe that these arguments against conformal gravity are quite solid. And as a theorist, I feel this is somewhat regrettable, because I find the theoretical foundations, the added conformal symmetry, and the quantum-field aspects of conformal gravity very appealing (and Mannheim's solution of the dark-matter problem is a nice bonus). But in its present form, conformal gravity does not look viable to me.

Acknowldedgment I thank Philip Mannheim for his patient explanations of various details of his theory. In discussions after this conference, Mannheim pointed out to me that the linear approximation I used in an earlier draft of this paper is not applicable to elementary particles, in which, he conjectures, the violent singularities in the energy-momentum tensor generate nonlinear gravitational effects. The extent to which nonlinearities come into play depends on the unknown internal structure of quarks, electrons, and whatever other elementary particles exist in atoms. To sidestep any dispute about the presence or absence of nonlinearities, I have rewritten the paper entirely in terms of Mannheim's nonlinear equations. For the arguments *i*) and *ii*) given here, my conclusions drawn from the linear and the nonlinear equations are the same. However, I wish to inform readers that Mannheim has not given his approval to my conclusions.

References

[1] P. D. Mannheim, see these Proceedings, Eqs. (56 – 58). Note that in my notation $dV' = 4\pi r'^2 dr'$, and note that because $B(r)$ appears on the right side, Eq. (3) is merely a formal, implicit solution.

[2] For bodies other than spheres, the $1/r$ potential in Eq. (3) acquires an angular dependence. For example, a source in the shape of a rod has a $1/r$ contribution in its potential at large distances perpendicular to the rod, but no $1/r$ contribution at all at any distance along the axis of the rod, in sharp contrast to the asymptotic omnidirectional Newtonian $1/r$ potential generated by the second-order Poisson equation.

[3] P. D. Mannheim, Progress in Particle and Nuclear Physics **56**, 340 (2006), Section 4.2

[4] C. Will, Theory and Experiment in Gravitational Physics (Cambridge University Press, Cambridge, 1993). The Eötvös data has been improved by a factor of about 3 since publication of Will's book, but the quoted number are more than adequate for present purposes.

[5] This can be seen more clearly in the linear approximation. The nuclear electric field is free of singularities and is sufficiently weak, so a linear approximation for Eq. (6) is valid, with $B(r) = 1$ on the left side of this equation. If, in contradiction to Mannheim's choice, we adopt a negative value of α_g , then the sign conflict in Eq. (6) can be avoided, but even so Eq. (6) cannot hold for every kind of nucleus, because it demands a universal proportionality between the integrals $\int r'^2 T_0^0 dV'$ and $\int T_0^0 dV'$. As we go through the Periodic Table, the ratio of these integrals changes by much more than the 4 parts in10^{10} permitted by the Eötvös data.

The Effect of Cosmic Inhomogeneities On The Average Cosmological Dynamics

T. P. Singh

Tata Institute of Fundamental Research,
Homi Bhabha Road, Mumbai 400 005, India.
email: tpsingh@tifr.res.in

Abstract

It is generally assumed that on sufficiently large scales the Universe is well-described as a homogeneous, isotropic FRW cosmology with a dark energy. Does the formation of nonlinear cosmic inhomogeneities produce a significant effect on the average large-scale FLRW dynamics? As an answer, we suggest that if the length scale at which homogeneity sets in is much smaller than the Hubble length scale, the back-reaction due to averaging over inhomogeneities is negligible. This result is supported by more than one approach to study of averaging in cosmology. Even if no single approach is sufficiently rigorous and compelling, they are all in agreement that the effect of averaging in the real Universe is small. On the other hand, it is perhaps fair to say that there is no definitive observational evidence yet that there indeed is a homogeneity scale which is much smaller than the Hubble scale, or for that matter, if today's Universe is indeed homogeneous on large scales. If the Copernican principle can be observationally established to hold, or is theoretically assumed to be valid, this provides strong evidence for homogeneity on large scales. However, even this by itself does not say what the scale of homogeneity is. If that scale is today comparable to the Hubble radius, only a fully non-perturbative analysis can establish or rule out the importance of cosmological back-reaction. This brief elementary report summarizes some recent theoretical developments on which the above inferences are based.

Based on a talk given at 'International Conference on Two Cosmological Models'
17-19 November, 2010, Mexico City, To appear in the Conference Proceedings

1 Introduction

The Universe that we see around us is lumpy - it has stars, galaxies, clusters of galaxies, superclusters, sheets, filaments and voids. We do not precisely know from observations what the size of the largest structures is; the size beyond which there are no larger structures. On the other hand the early Universe is very well-described as a homogeneous, isotropic FRW cosmology [Big-Bang nucleosynthesis and the relic CMB are evidence of success] and the present Universe is well-described as an FRW cosmology with dark energy. How does one reconcile a universe which is observed to be inhomogeneous and anisotropic on smaller scales, with a universe that is assumed to be homogeneous and isotropic on large scales? Clearly, some way of averaging the matter distribution and the related Einstein equations has to be invoked. What is the right way? The true metric of the universe is the one produced by the inhomogeneous matter distribution. On large scales, one assumes the average of the true metric to be FLRW, constructs the Einstein tensor for it, and uses it on the left hand side of the Friedmann equations wherein the matter content on the right hand side is a perfect fluid. Because Einstein equations are nonlinear, the Einstein tensor constructed from the average metric tensor will in general not be the same as the average of the Einstein tensor of the true metric :

$$< G_{\mu\nu}(g_{\mu\nu}) > = < T_{\mu\nu} > \neq \ G_{\mu\nu}(< g_{\mu\nu} >) \tag{1}$$

Here $< ... >$ denotes the averaging operation [whose correct definition, for tensors on a curved spacetime, is itself a major challenge]; and $< g_{\mu\nu} >= g_{\mu\nu}|_{FLRW}$. The correct averaged Einstein equations are of course given by the first pair (the equality) in the above set, whereas in cosmology we assume the correct equations to be those given by the second pair (by assuming the inequality to actually be an equality). This is obviously done because the latter option is infinitely simpler - it is straightforward to write the Einstein tensor for the Robertson-Walker metric, but it is impossible to find the true metric of the inhomogeneous universe and then average its corresponding Einstein tensor. Since the first and the third terms in (1) could differ significantly, we might be working with the wrong averaged Einstein equation on cosmological scales. This is the averaging problem: how to correctly average Einstein equations, and to find out if the neglected terms (the so-called back-reaction) can become important in the late stages of an evolving universe, when nonlinear structures such as galaxies and clusters form. In particular, can the back-reaction mimic a dark energy, and explain the observed cosmic acceleration?

The problem of averaging of Einstein equations has a long history, and has recently been reviewed in an important article by Ellis [1]. Important contributions to the study of averaging have been made in recent times, amongst others, by Buchert [2], Coley [3], Wald [4], Zalaletdinov [5] and their collaborators. Specific applications have been developed by Kolb [6], Marra [7], Rasanen [8], Sussman [9], Wiltshire [10] and others. Much of this work, as well as earlier developments, are reviewed by Ellis, and we will not enter into details here, except in the context of specific arguments developed here.

If we want to find out the back-reaction on an FLRW universe, it certainly means we are taking an FLRW geometry as given on large scales. It is hence necessary to first know what the observational evidence for large scale homogeneity and isotropy is, and what is the length scale at which homogeneity sets in.

2 Evidence for large scale homogeneity

Careful discussions of this issue have recently been given by Clarkson and Maartens [11], Maartens [12] and Ellis [1]. What we have to say below is a summary from these earler works, and is reported here because of its significance for the discussion on averaging in the next section.

It is well known that homogeneity on spatial surfaces can not be established by direct observations, because all our observations are on the past light-cone. Hence tests of homogeneity have to exploit the following route : isotropy around us along with the Copernican Principle [CP] implies homogeneity, and hence a FLRW Universe. So one needs to test for isotropy and for CP independently.

The observational evidence for spacetime isotropy around our world-line can be investigated from examining the isotropy of the CMB and the galaxy distribution. For a perfectly isotropic CMB, all multipoles of the distribution function higher than the monopole, as well their time derivatives, vanish. However without CP one cannot deduce the vanishing of the spatial derivatives of the higher multipoles, and hence spacetime isotropy about our world-line cannot be deduced without CP. As for baryonic matter (along with certain assumptions for the distribution of CDM and dark energy) it can be shown that isotropic distribution of the following four matter observables on the light-cone implies an isotropic spacetime geometry : angular diameter distances, galaxy number counts, bulk velocities and lensing (details and references to original work can be found in [12]). As pointed out by Maartens, it is not known whether almost-isotropy of observations leads to almost-isotropy of spacetime geometry.

Next, one considers what can be inferred about spatial homogeneity, if one assumes CP, and considers the following three cases : isotropic matter distribution, isotropic CMB, almost-isotropic CMB. If all fundamental observers measure the same isotropic distribution of the four matter observables mentioned above, this implies homogeneity, and the Universe is FLRW. It can be proved that exact isotropy of the CMB for all observers also implies an FLRW universe. Almost-isotropy of the CMB can be shown, via a non-perturbative analysis, to imply an almost-FLRW universe, provided some of the time and spatial derivatives of the multipoles are sufficiently small.

Thus it is clear that the case for an almost-FLRW universe will be strong if observational tests support the Copernican Principle. These tests can be carried out by testing the standard consistency relations in FLRW geometry. The FLRW curvature parameter which can be inferred from geometric measurements is independent of redshift, and a detection of redshift dependence of this parameter will indicate departure from homogeneity. A second test is the time drift of cosmological redshift, and a third test is a significant difference between the radial and transverse BAO scales. None of these tests have yet been carried out, but their eventual execution will play a crucial role in confirming large-scale homogeneity. The CP can also be tested by looking for a large thermal or kinetic Sunyaev-Zeldovich effect temeperature distortion of the CMB. Also, a large SZ effect induced CMB polarization could indicate a violation of CP and hence of homogeneity.

As of now, there is no evidence against CP, but neither is there clinching evidence for large-scale homogeneity. Also, it is not quite clear at what scale homogeneity sets in. If we assume that there is homogeneity, and that too at a scale much less than the Hubble scale, say at around 100 MPc, then it can be shown (as discussed next) that the cosmological back-reaction is negligible. And the ΛCDM model is then a good description of the present day Universe. On the other hand if there are much larger nonlinear structures in the Universe - their formation can then no longer be described perturbatively on an FLRW background, and the back-reaction problem will have to be examined afresh.

3 Averaging in Cosmology and Calculation of Back-Reaction

Assuming that the scale at which homogeneity sets in is much smaller than the Hubble scale, we give three explanations as to why the back-reaction will be small : (i) a simple argument due to Peebles [13]; (ii) our own work [14], [15], [16], [17] which builds on Zalaletdinov's Macroscopic Theory of Gravity [MG] [18], [19], [20], [21]; and (iii) the work of Wald and collaborators [22], [4]. Similar results obtained by a few other researchers, which support the present inference, are briefly reviewed in Paranjape's thesis [23].

3.1 An argument due to Peebles ([13] and references therein)

For nonlinear structures such as galaxies, the Newtonian gravitational potential is of the order of the square of the velocity dispersion [about 300 km/sec], i.e. $\phi \sim 10^{-6}$. Hence the galaxy distribution can be described as a perturbation over an FLRW universe. The metric can be written as a perturbed FLRW Universe and the Einstein equations can be split into an evolution equation for the background scale-factor and the Poisson equation for the perturbed Newtonian potential determined by the density contrast (assumed to be provided by non-interacting dark matter).

In order to find the effect of averaging on the FLRW equations, spatial averages of Einstein equations need to be computed to order ϕ^2, in particular for the dominant term which is proportional to $\nabla\phi.\nabla\phi$. When this is done, one finds corrections due to back-reaction in both the Friedmann equations - corrections in the form of a kinetic energy coming from the mean square velocity dispersion, and the averaged gravitational potential energy determined by the density contrast of the formed nonlinear structures. Both these correction terms are of the order of a part in a million, and hence much smaller than the magnitude of the observed dark energy.

The discussion by Peebles is patterned in part on the nice work of Siegel and Fry [24]. It seems to us that there is room for improvement in this argument : one should not fix the background, but allow for the possibility that as perturbations grow, the background around which back-reaction should be calculated may itself be changing, because of feedback from the perturbations. One has to ascertain that a runaway process leading to breakdown of perturbation theory does not take place. This is the study we attempted by applying Zalaletdinov's averaging theory [Macroscopic Gravity] to cosmology. Before we summarize our work on applying MG, it will be useful to briefly review Buchert's averaging scheme. We do this because the Buchert approach provides simple averaged equations, while being less ambitious than MG. Remarkably, the averaged equations that arise from MG are very similar to Buchert's averaging equations, enforcing a certain high degree of reliability of both approaches, despite their conceptual differences.

3.2 Buchert's averaging scheme for a dust spacetime

For a general spacetime containing irrotational dust, the metric can be written as

$$ds^2 = -dt^2 + h_{ij}(\vec{x}, t)dx^i dx^j \,. \tag{2}$$

The expansion tensor Θ^i_j is given by $\Theta^i_j \equiv (1/2)h^{ik}\dot{h}_{kj}$ where the dot refers to a derivative with respect to time t. The traceless symmetric shear tensor is defined as $\sigma^i_j \equiv \Theta^i_j - (\Theta/3)\delta^i_j$ where $\Theta = \Theta^i_i$ is the expansion scalar. The Einstein equations can be split into a set of scalar equations and a set of vector and traceless tensor equations. The scalar equations are the Hamiltonian constraint (3a)

and the evolution equation for Θ (3b),

$$ {}^{(3)}\mathcal{R} + \frac{2}{3}\Theta^2 - 2\sigma^2 = 16\pi G\rho \tag{3a} $$

$$ {}^{(3)}\mathcal{R} + \dot{\Theta} + \Theta^2 = 12\pi G\rho \tag{3b} $$

where the dot denotes derivative with respect to time t, ${}^{(3)}\mathcal{R}$ is the Ricci scalar of the 3-dimensional hypersurface of constant t and σ^2 is the rate of shear defined by $\sigma^2 \equiv (1/2)\sigma^i_j\sigma^j_i$. Eqns. (3a) and (3b) can be combined to give Raychaudhuri's equation

$$ \dot{\Theta} + \frac{1}{3}\Theta^2 + 2\sigma^2 + 4\pi G\rho = 0\,. \tag{4} $$

The continuity equation $\dot{\rho} = -\Theta\rho$ which gives the evolution of ρ, is consistent with Eqns. (3a), (3b). We only consider the scalar equations, since the spatial average of a scalar quantity can be defined in a gauge covariant manner within a given foliation of space-time. For the space-time described by (2), the spatial average of a scalar $\Psi(t,\vec{x})$ over a *comoving* domain $\mathcal{D}$ at time t is defined by

$$ \langle\Psi\rangle = \frac{1}{V_\mathcal{D}}\int_\mathcal{D} d^3x\sqrt{h}\,\Psi \tag{5} $$

where h is the determinant of the 3-metric h_{ij} and $V_\mathcal{D}$ is the volume of the comoving domain given by $V_\mathcal{D} = \int_\mathcal{D} d^3x\sqrt{h}$.

Spatial averaging is, by definition, not generally covariant. Thus the choice of foliation is relevant, and should be motivated on physical grounds. In the context of cosmology, averaging over freely-falling observers is a natural choice, especially when one intends to compare the results with standard FRW cosmology. Following the definition (5) the following commutation relation then holds [2]

$$ \langle\Psi\rangle^{\cdot} - \langle\dot{\Psi}\rangle = \langle\Psi\Theta\rangle - \langle\Psi\rangle\langle\Theta\rangle \tag{6} $$

which yields for the expansion scalar Θ

$$ \langle\Theta\rangle^{\cdot} - \langle\dot{\Theta}\rangle = \langle\Theta^2\rangle - \langle\Theta\rangle^2\,. \tag{7} $$

Introducing the dimensionless scale factor $a_\mathcal{D} \equiv (V_\mathcal{D}/V_{\mathcal{D}in})^{1/3}$ normalized by the volume of the domain $\mathcal{D}$ at some initial time t_{in}, we can average the scalar Einstein equations (3a), (3b) and the continuity equation to obtain

$$ \langle\Theta\rangle = 3\frac{\dot{a}_\mathcal{D}}{a_\mathcal{D}}\,, \tag{8a} $$

$$ 3\left(\frac{\dot{a}_\mathcal{D}}{a_\mathcal{D}}\right)^2 - 8\pi G\langle\rho\rangle + \frac{1}{2}\langle\mathcal{R}\rangle = -\frac{\mathcal{Q}_\mathcal{D}}{2}\,, \tag{8b} $$

$$ 3\left(\frac{\ddot{a}_\mathcal{D}}{a_\mathcal{D}}\right) + 4\pi G\langle\rho\rangle = \mathcal{Q}_\mathcal{D}\,, \tag{8c} $$

$$ \langle\rho\rangle^{\cdot} = -\langle\Theta\rangle\langle\rho\rangle = -3\frac{\dot{a}_\mathcal{D}}{a_\mathcal{D}}\langle\rho\rangle\,. \tag{8d} $$

Here $\langle\mathcal{R}\rangle$, the average of the spatial Ricci scalar ${}^{(3)}\mathcal{R}$, is a domain dependent spatial constant. The 'backreaction' $\mathcal{Q}_\mathcal{D}$ is given by

$$\mathcal{Q}_\mathcal{D} \equiv \frac{2}{3}\left(\langle\Theta^2\rangle - \langle\Theta\rangle^2\right) - 2\langle\sigma^2\rangle \tag{9}$$

and is also a spatial constant. The last equation (8d) simply reflects the fact that the mass contained in a comoving domain is constant by construction : the local continuity equation $\dot{\rho} = -\Theta\rho$ can be solved to give $\rho\sqrt{h} = \rho_0\sqrt{h_0}$ where the subscript 0 refers to some arbitrary reference time t_0. The mass $M_\mathcal{D}$ contained in a comoving domain $\mathcal{D}$ is then $M_\mathcal{D} = \int_\mathcal{D} \rho\sqrt{h}d^3x = \int_\mathcal{D} \rho_0\sqrt{h_0}d^3x = \text{constant}$. Hence

$$\langle\rho\rangle = \frac{M_\mathcal{D}}{V_{\mathcal{D}in}a_\mathcal{D}^3} \tag{10}$$

which is precisely what is implied by Eqn. (8d).

This averaging procedure can only be applied for spatial scalars, and hence only a subset of the Einstein equations can be smoothed out. As a result it may appear that the outcome of such an approach is severely restricted, and essentially incomplete due to the impossibility to analyze the full set of equations. However one should note that the cosmological parameters of interest are scalars, and the averaging of the exact scalar part of Einstein equations provides the requisite needed information. A more general strategy would be to consider the smoothing of tensors, which is beyond the scalar approach that certainly provides useful information, albeit not the full information.

The dynamical equations above can be cast in a form which is immediately comparable with the standard FRW equations [2]. Namely,

$$\frac{\ddot{a}_\mathcal{D}}{a_\mathcal{D}} = -\frac{4\pi G}{3}\left(\rho_{\text{eff}} + 3P_{\text{eff}}\right) \tag{11a}$$

$$\left(\frac{\dot{a}_\mathcal{D}}{a_\mathcal{D}}\right)^2 = \frac{8\pi G}{3}\rho_{\text{eff}} \tag{11b}$$

with ρ_{eff} and P_{eff} defined as

$$\rho_{\text{eff}} = \langle\rho\rangle - \frac{\mathcal{Q}_\mathcal{D}}{16\pi G} - \frac{\langle\mathcal{R}\rangle}{16\pi G} \quad ; \quad P_{\text{eff}} = -\frac{\mathcal{Q}_\mathcal{D}}{16\pi G} + \frac{\langle\mathcal{R}\rangle}{48\pi G}\,. \tag{12}$$

A necessary condition for (11a) to integrate to (11b) takes the form of the following differential equation involving $\mathcal{Q}_\mathcal{D}$ and $\langle\mathcal{R}\rangle$

$$\dot{\mathcal{Q}}_\mathcal{D} + 6\frac{\dot{a}_\mathcal{D}}{a_\mathcal{D}}\mathcal{Q}_\mathcal{D} + \langle\mathcal{R}\rangle^{\cdot} + 2\frac{\dot{a}_\mathcal{D}}{a_\mathcal{D}}\langle\mathcal{R}\rangle = 0 \tag{13}$$

and the criterion to be met in order for the effective scale factor $a_\mathcal{D}$ to accelerate, is

$$\mathcal{Q}_\mathcal{D} > 4\pi G\langle\rho\rangle\,. \tag{14}$$

The Buchert scheme has been applied extensively, and in particular can be used to show that there indeed are toy cosmological models which when averaged over inhomogeneities can produce an apparent acceleration. However, not all Einstein equations are averaged, and one does not have an averaged metric here [which we would like to be the FLRW metric]. Macroscopic Gravity can achieve that, while reproducing modified Friedmann equtions analogous to the Buchert equations, when applied to cosmology.

4 Macroscopic Gravity

This theory is developed comprehensively in the works of Zalaletdinov; briefly introduced in [14], and reviewed in Paranjape's thesis [23]. For detailed discussions and the primary interpretation of MG, the reader is referred to Zalaletdinov's original papers cited in this article.

For the purpose of averaging of tensors the key new element which is introduced is a *bivector* $\mathcal{W}_b^{a'}(x', x)$ which transforms as a vector at event x' and as a co-vector at event x. The bivector is used to define the "bilocal extension" of a general tensorial object

$$\widetilde{P}^a(x', x) = \mathcal{W}^a_{a'}(x, x')P^{a'}(x') \tag{15}$$

The "average" of $P^a(x)$ over a 4-dimensional spacetime region $\mathbf{\Sigma}$ with a supporting point x is

$$\bar{P}^a(x) = \left\langle \widetilde{P}^a \right\rangle_{ST} = \frac{1}{V_\Sigma} \int_\Sigma d^4x' \sqrt{-g'} \widetilde{P}^a(x', x) \tag{16}$$

and this averaging operation preserves tensorial properties.

There is a certain degree of non-uniqueness in te choice of the coordination bi-vector - the freedom coming from the presence of undetermined structure constants in the commutation relations for a vector basis in terms of which one can solve for the cordination bivector. The simplest choice is to set these structure constants to zero. When that is done, then in a volume preserving coordinate system ϕ^m, [VPC], i.e. one with $g(\phi^m) =$ constant, the coordination bivector takes its most simple form, namely

$$\mathcal{W}_j^{a'}(x', x) \mid_{\text{proper}} = \delta_j^{a'} \,. \tag{17}$$

The effect of this non-uniqueness on the physical results for averaging in cosmology remains to be estimated. Nonetheless, it is noteworthy that the averaged Friedmann equations to be derived from this approach are similar to Buchert's and the physical results about the magnitude of the back-reaction is identical to the one due to Peebles. This gives confidence in the robustness of the results obtained, even though there is freedom in the choice of the coordination bivector. It is also useful to note that this bi-vector is different from the Synge bi-tensor which leaves the metric invariant upon avaeraging, and hence cannot really be used to average an inhomogeneous geometry.

Averaged Geometry : the key idea of Macroscopic Gravity is that the average connection $\bar{\mathbf{\Omega}}^a{}_b(x)$

$$\bar{\mathbf{\Omega}}^a{}_b \equiv \langle \mathbf{\Omega^a_b} \rangle \,, \tag{18}$$

is defined as the connection 1-form on a new, averaged manifold $\bar{\mathcal{M}}$. Next one defines a correlation 2-form

$$\mathbf{Z}^{a\ i}_{\ b\ j} = \left\langle \mathbf{\Omega^a_b} \wedge \mathbf{\Omega^i_j} \right\rangle_{ST} - \bar{\mathbf{\Omega}}^a{}_b \wedge \bar{\mathbf{\Omega}}^i{}_j \,. \tag{19}$$

Denoting $\mathbf{R}^a{}_b \equiv \langle \widetilde{\mathbf{r}}^a{}_b \rangle_{ST}$, where $\mathbf{r}^a_b$ is the curvature 2-form of the inhomogeneous geometry, and the curvature 2-form on the averaged manifold $\bar{\mathcal{M}}$ as $\mathbf{M}^a{}_b$ can be shown to give

$$\mathbf{M}^a{}_b = \mathbf{R}^a{}_b - \mathbf{Z}^{a\ c}_{\ c\ b} \,. \tag{20}$$

The inhomogeneous Einstein equations

$$g^{ak} r_{kb} - \frac{1}{2} \delta^a_b g^{ij} r_{ij} = -\kappa t^{a(\text{mic})}_b \,, \tag{21}$$

average out to

$$E^a_b = -\kappa T^a_b + C^a_b \,, \tag{22}$$

$$C^a_b = \left(Z^a{}_{ijb} - \frac{1}{2}\delta^a_b Z^m{}_{ijm} \right) G^{ij}. \tag{23}$$

G^{ij} is the metric on the averaged geometry. The correlation 2-form is assumed to satisfy certain differential conditions which amount to closure conditions for the above system of averaged equations.

These averaged equations of Macroscopic Gravity carry, in a covariant and non-perturbative manner, information about the effect of the underlying inhomogeneities on the averaged geometry.

5 Application of Macroscopic Gravity to Cosmology

In order to apply MG to cosmology we start with the assumption that Einstein's equations are to be imposed on length scales where stars are pointlike objects (we denote such a scale as $L_{\rm inhom}$). The averaging we perform will be directly at a length scale $L_{\rm FLRW}$ larger than about $100h^{-1}$Mpc or so. This averaging scale is assumed to satisfy $L_{\rm inhom} \ll L_{\rm FLRW} \ll L_{\rm Hubble}$ where $L_{\rm Hubble}$ is the length scale of the observable universe. The averaging will be assumed to yield a geometry which has homogeneous and isotropic spatial sections. In other words, we will assume that the averaged manifold $\bar{\mathcal{M}}$ admits a preferred, hypersurface-orthogonal unit timelike vector field $\bar{v}^a$, which defines 3-dimensional spacelike hypersurfaces of constant curvature, and that $\bar{v}^a$ is tangent to the trajectories of observers who see an isotropic Cosmic Background Radiation. (These "observers" are defined in the averaged manifold – we will clarify below what they correspond to in the inhomogeneous manifold.) Throughout the rest of this article, for simplicity, we will work with the special case where the spatial sections on $\bar{\mathcal{M}}$ defined by $\bar{v}^a$ are flat. (In principle the entire calculation can be repeated for non-flat spatial sections as well.) One can then choose coordinates (t, x^A), $A = 1, 2, 3$, on $\bar{\mathcal{M}}$ such that the spatial line element takes the form

$${}^{(\bar{\mathcal{M}})}ds^2_{\rm spatial} = a^2(t)\delta_{AB}dx^A dx^B \,, \tag{24}$$

where $\delta_{AB} = 1$ for $A = B$, and 0 otherwise, and we have $\bar{v}^a = (\bar{v}^t, 0, 0, 0)$ so that the spatial coordinates are comoving with the preferred observers. The vector field $\bar{v}^a$ also defines a proper time (the cosmic time) τ such that $\partial_\tau = \bar{v}^a\partial_a = \bar{v}^t\partial_t$. We will further assume that the averaged energy-momentum tensor T^a_b can be written in the form of a perfect fluid, as

$$T^a_b = \rho \bar{v}^a \bar{v}_b + p\pi^a_b \,, \tag{25}$$

where the projection operator π^a_b is defined as

$$\pi^a_b = \delta^a_b + \bar{v}^a\bar{v}_b \,, \tag{26}$$

and ρ and p are the homogeneous energy density and pressure respectively, as measured by observers moving on trajectories (in $\bar{\mathcal{M}}$) with the tangent vector field $\bar{v}^a$,

$$\rho \equiv T^a_b \bar{v}^b \bar{v}_a \quad ; \quad p \equiv \frac{1}{3}\pi^b_a T^a_b \,. \tag{27}$$

ρ and p are observationally relevant quantities, since all measurements of the matter energy density, especially those from studies of Large Scale Structure, interpret observations in the context of the averaged geometry. An important consequence of the above assumptions is that the correlation tensor C^a_b, when expressed in terms of the natural coordinates adapted to the spatial sections defined by the vector field $\bar{v}^a$, is *spatially homogeneous*. This is clear when the modified Einstein equations are written in these natural coordinates.

The existence of the vector field $\bar{v}^a$ with the attendant assumptions described above, allows us to separate out the nontrivial components of the (FLRW) Einstein tensor E^a_b on $\bar{\mathcal{M}}$ in a coordinate independent fashion – the Einstein tensor can be written as

$$E^a_b = j_1(x)\bar{v}^a\bar{v}_b + j_2(x)\pi^a_b$$
$$j_1(x) \equiv E^a_b\bar{v}^b\bar{v}_a \;\; ; \;\; j_2(x) \equiv \frac{1}{3}\left(\pi^b_a E^a_b\right) \, , \tag{28}$$

where $j_1(x)$ and $j_2(x)$ are scalar functions whose form depends upon the coordinates used. The remaining components given by $\pi^b_k E^a_b\bar{v}_a$ and the traceless part of $\pi^i_a\pi^b_k E^a_b$, vanish identically. Since the energy-momentum tensor T^a_b in Eqn. (25) also has an identical structure, this structure is therefore *also imposed* on the correlation tensor C^a_b. Namely, $\pi^b_k C^a_b\bar{v}_a$ and the traceless part of $\pi^i_a\pi^b_k C^a_b$ *must vanish*. This is a condition on the underlying inhomogeneous geometry, irrespective of the coordinates used on either $\mathcal{M}$ or $\bar{\mathcal{M}}$, and is clearly a consequence of demanding that the averaged geometry have the symmetries of the FLRW spacetime.

This leads us to the crucial question of the choice of *gauge* for the underlying geometry : namely, what choice of spatial sections for the *inhomogeneous* geometry, will lead to the spatial sections of the FLRW metric in the comoving coordinates defined in Eqn. (24)? Since the matter distribution at scale $L_{\rm inhom}$ need not be pressure-free (or, indeed, even of the perfect fluid form), there is clearly no natural choice of gauge available, although locally, a synchronous reference frame can always be chosen. We note that there must be *at least one* choice of gauge in which the averaged metric has spatial sections in the form (24) – this is simply a refinement of the Cosmological Principle, and of the Weyl postulate, according to which the Universe is homogeneous and isotropic on large scales, and individual galaxies are considered as the "observers" travelling on trajectories with tangent $\bar{v}^a$. In the averaging approach, it makes more sense to replace "individual galaxies" with the *averaging domains* considered as physically infinitesimal cells – the "points" of the averaged manifold $\bar{\mathcal{M}}$. This is physically reasonable since we know after all, that individual galaxies exhibit peculiar motions, undergo mergers and so on. This idea is also more in keeping with the notion that the Universe is homogeneous and isotropic *only on the largest scales*, which are much larger than the scale of individual galaxies.

Consider any $3+1$ spacetime splitting in the form of a lapse function $N(t, x^J)$, a shift vector $N^A(t, x^J)$, and a metric for the 3-geometry $h_{AB}(t, x^J)$, so that the line element on $\mathcal{M}$ can be written as

$$^{(\mathcal{M})}ds^2 = -\left(N^2 - N_A N^A\right)dt^2 + 2N_B dx^B dt + h_{AB}dx^A dx^B \, , \tag{29}$$

where $N_A = h_{AB}N^B$. At first sight, it might seem reasonable to leave the choice of gauge arbitrary. However the analysis is then complicated. On the other hand, if we make the assumption that the spatial sections on $\mathcal{M}$ leading to the spatial metric (24) on $\bar{\mathcal{M}}$, are spatial sections *in a volume preserving gauge*, then the correlation terms simplify greatly. This is not surprising since the MG formalism is nicely adapted to the choice of volume preserving coordinates. The case when the gauge is left unspecified is dealt with in our original papers.

To begin our calculation, we perform a coordinate transformation and shift to the gauge wherein the new lapse function N is given by $N = 1/\sqrt{h}$ where h is the determinant of the new 3-metric h_{AB}. In general, one will now be left with a non-zero shift vector N^A; however, the condition $N\sqrt{h} = 1$ ensures that the coordinates we are now using are volume preserving, since the metric determinant is given by $g = -N^2 h = -1 = \text{constant}$. We denote these volume preserving coordinates (VPCs) by $(\bar{t}, \mathbf{x}) = (\bar{t}, x^A) = (\bar{t}, x, y, z)$, and will assume that the spatial coordinates are non-compact. For simplicity, we make the added assumption that $N^A = 0$ in the inhomogenous geometry, so that $g_{\bar{t}\bar{t}} = -N^2 = -1/h$ and $g_{\bar{t}A} = 0$. The line element for the inhomogenous manifold $\mathcal{M}$ becomes

$$ {}^{(\mathcal{M})}ds^2 = -\frac{d\bar{t}^2}{h(\bar{t}, \mathbf{x})} + h_{AB}(\bar{t}, \mathbf{x}) dx^A dx^B \, . \tag{30} $$

Note that in this gauge, the average takes on a particularly simple form : for a tensor $p^i_j(x)$, with a spacetime averaging domain given by the "cuboid" Σ defined by

$$ \Sigma = \{(\bar{t}, x, y, z) \mid -T/2 < \bar{t} < T/2, -L/2 < x, y, z < L/2\} \, , \tag{31} $$

where T and L are averaging time and length scales respectively, the average is given by

$$ \begin{aligned} &\left\langle \widetilde{p}^{\,i}_j \right\rangle_{ST} (\bar{t}, \mathbf{x}) = \left\langle p^i_j \right\rangle_{ST} (\bar{t}, \mathbf{x}) \\ &= \frac{1}{TL^3} \int_{\bar{t}-T/2}^{\bar{t}+T/2} dt' \int_{-L/2}^{+L/2} dx' dy' dz' \left[p^i_j(t', x', y', z') \right] , \end{aligned} \tag{32} $$

where the limits on the spatial integral are understood to hold for all three spatial coordinates. We define the "spatial averaging limit" as the limit $T \to 0$ (or $T \ll L_{\text{Hubble}}$) which is interpreted as providing a definition of the average on a spatial domain corresponding to a "thin" time slice, the averaging operation now being given by

$$ \begin{aligned} &\langle p^i_j \rangle (\bar{t}, \mathbf{x}) \\ &= \frac{1}{L^3} \int_{-L/2}^{+L/2} dx' dy' dz' \left[p^i_j(\bar{t}, x', y', z') \right] + \mathcal{O}\left(T L^{-1}_{\text{Hubble}} \right) \, . \end{aligned} \tag{33} $$

(Note the time dependence of the integrand.) Henceforth, averaging will refer to spatial averaging, and will be denoted by $\langle ... \rangle$, in contrast to the spacetime averaging considered thus far (denoted by $\langle ... \rangle_{ST}$). The choice of a cube with sides of length L as the spatial averaging domain was arbitrary, and is in fact not essential for any of the calculations to follow. In particular, all calculations can be performed with a spatial domain of arbitrary shape. We will only use the cube for definiteness and simplicity in displaying equations. The significance of introducing a spatial averaging in this manner is that the construction of spatial averaging is not isolated from spacetime averaging, but is a special limiting case of the latter and is, in fact, still a fully covariant operation.

For the volume preserving gauge, we have

$$ \begin{aligned} G_{\bar{t}\bar{t}} = \langle g_{\bar{t}\bar{t}} \rangle = \langle \frac{-1}{h} \rangle = -f^2(\bar{t}) \, ; \\ G_{AB} = \langle h_{AB} \rangle = \bar{a}^2(\bar{t}) \delta_{AB} \, , \end{aligned} \tag{34} $$

where $\bar{a}$ and f are some functions of the time coordinate alone. A few remarks are in order on this particular choice of assumptions. Apart from the fact that the spacetime averaging operation takes

on its simplest possible form (32) in this gauge and allows a transparent definition of the spatial averaging limit, it can also be shown that the assumptions in Eqn. (34) are sufficient to establish the following relations :

$$f^2(\bar{t}) = \langle \frac{1}{h} \rangle = \frac{1}{\langle h \rangle} = \frac{1}{\bar{a}^6} \,. \tag{35}$$

Here the second equality arises from the condition $\bar{g}^{ij} = G^{ij}$ which can be assumed whenever the averaged metric is of the FLRW form. The last equality follows on considering the conditions $\langle \widetilde{\Gamma}^a_{bc} \rangle = {}^{(\mathrm{FLRW})}\Gamma^a_{bc}$ in obvious notation, (the basic assumption of the MG averaging scheme). Eqn. (35) reduces the line element on $\bar{\mathcal{M}}$ to the form

$${}^{(\bar{\mathcal{M}})}ds^2 = -\frac{d\bar{t}^2}{\bar{a}^6(\bar{t})} + \bar{a}^2(\bar{t})\delta_{AB}dx^A dx^B \ . \tag{36}$$

The line element in Eqn. (36) clearly corresponds to the FLRW metric in a *volume preserving gauge*. In other words, the (spatial) average of the inhomogeneous geometry in the volume preserving gauge leads to a geometry with homogeneous and isotropic spatial sections, also in a volume preserving gauge. Note that the gauge in Eqn. (36) for the FLRW spacetime differs from the standard synchronous and comoving gauge, only by a redefinition of the time coordinate. The vector field $\bar{v}^a$ introduced at the beginning of this section and which defines the FLRW spatial sections, is now given by

$$\bar{v}^a = \left(\bar{a}^3, 0, 0, 0\right) \quad ; \quad \bar{v}_a = G_{ab}\bar{v}^b = \left(-\frac{1}{\bar{a}^3}, 0, 0, 0, \right) \,. \tag{37}$$

Before proceeding to the calculation of the correlation terms and the averaged Einstein equations, we briefly describe why it is important to consider the spatial averaging limit of the MG averaging operation. The key idea to emphasize is that an average of the homogeneous and isotropic FLRW geometry, should give back the same geometry. Since the FLRW geometry has a preferred set of spatial sections, it is important therefore to perform the averaging over these sections. Further, since the FLRW metric adapted to its preferred spatial sections depends on the time coordinate, it is also essential that the spacetime average should involve a time range that is short compared to the scale over which say the scale factor changes significantly. Clearly then, averaging the FLRW metric (denoted ${}^{(FLRW)}g_{ab}$) given in Eqn. (36) (which is in volume preserving gauge) will strictly yield the same metric *only* in the limit $T \to 0$. Namely, for the cuboid Σ defined in Eqn. (31)

$$\begin{aligned} \langle {}^{(FLRW)}\widetilde{g}_{ab} \rangle &= \lim_{T\to 0} \frac{1}{TL^3} \int_\Sigma dt' d^3x' \, {}^{(FLRW)}g_{ab}(t', \mathbf{x}') \\ &= {}^{(FLRW)}g_{ab} \,, \end{aligned} \tag{38}$$

which should be clear from the definition of the metric. The result $\langle {}^{(FLRW)}\widetilde{g}_{ab} \rangle = {}^{(FLRW)}g_{ab}$ in the spatial averaging limit can also be shown to hold for the FLRW metric in synchronous gauge, where the coordination bivector $\mathcal{W}^{a'}_j$ can be easily computed using the transformation from the VPCs $(\bar{t}, x^A)$ to the synchronous coordinates (τ, y^A) given by

$$\tau = \int^{\bar{t}} \frac{dt}{\bar{a}^3(t)} \quad ; \quad y^A = x^A \,. \tag{39}$$

The transformation (39) will also later allow us to write the averaged equations in the synchronous gauge for the averaged geometry.

We now proceed to calculating the correlation 2-form $\mathbf{Z}^{a\,i}_{b\,j}$ and thereby the averaged Einstein equations.

6 The averaged cosmological field equations

We start by defining (in any gauge with $N^A = 0$) the expansion tensor Θ^A_B by

$$\Theta^A_B \equiv \frac{1}{2N} h^{AC} \dot{h}_{CB} \,, \tag{40}$$

where the dot will always refer to a derivative with respect to the VPC time $\bar{t}$, and h^{AB} is the inverse of the 3-metric h_{AB}. (This also gives the symmetric tensor $\Theta_{AB} = (1/2N)\dot{h}_{AB}$, which is the negative of the extrinsic curvature tensor.) The traceless symmetric shear tensor σ^A_B and the shear scalar σ^2 are defined by

$$\sigma^A_B \equiv \Theta^A_B - (\Theta/3)\delta^A_B \;\;;\;\; \sigma^2 \equiv \frac{1}{2}\sigma^A_B \sigma^B_A \,, \tag{41}$$

where $\Theta \equiv \Theta^A_A = (1/N)\partial_{\bar{t}} \ln \sqrt{h}$ is the expansion scalar.

The connection 1-forms $\omega^i{}_j = \Gamma^i_{jk}\mathbf{d}x^k$ can be easily calculated in terms of the expansion tensor, for an arbitrary lapse function N. Specializing to the volume preserving gauge ($N = h^{-1/2}$), the bilocal extensions $\mathbf{\Omega^i_j}$ of the connection 1-forms are trivial and are simply given by

$$\mathbf{\Omega^i_j}(x', x) = \Gamma^i_{jk}(x')\mathbf{d}x^k \,. \tag{42}$$

Since $G_{ab} = \bar{g}_{ab}$, the connection 1-forms $\bar{\mathbf{\Omega}}^i{}_j$ for the averaged manifold $\bar{\mathcal{M}}$ are constructed using the FLRW metric in volume preserving gauge given in Eqn. (36), and can also be easily evaluated.

We can now construct the correlation 2-form $\mathbf{Z}^{a\,i}_{b\,j}$ and from there the correlation tensor :

$$C^a_b = \left(Z^a{}_{ijb} - \frac{1}{2}\delta^a_b Z^m{}_{ijm} \right) G^{ij}. \tag{43}$$

Now, the components of the Einstein tensor E^a_b for the averaged spacetime with metric (36) are given by

$$\begin{aligned} E^{\bar{t}}_{\bar{t}} &= 3\bar{a}^6 H^2 \;\;;\;\; E^{\bar{t}}_A = 0 = E^B_{\bar{t}} \,, \\ E^A_B &= \bar{a}^6 \delta^A_B \left[2\left(\frac{\ddot{\bar{a}}}{\bar{a}} + 3H^2 \right) + H^2 \right] \,, \end{aligned} \tag{44}$$

where the peculiar splitting of terms in the last equation is for later convenience. Recall that the overdot denotes a derivative with respect to the VPC time $\bar{t}$, not synchronous time. In terms of the coordinate independent objects introduced in Eqn. (28), we have

$$j_1(x) = -3\bar{a}^6 H^2 \;\;;\;\; j_2(x) = \bar{a}^6 \left[2\left(\frac{\ddot{\bar{a}}}{\bar{a}} + 3H^2 \right) + H^2 \right] . \tag{45}$$

From the averaged Einstein equations we next construct the scalar equations which in the standard case would correspond to the Friedmann equation and the Raychaudhuri equation. These correspond to the Einstein tensor components,

$$E^a_b \bar{v}^b \bar{v}_a = j_1(x) \;\;;\;\; \pi^b_a E^a_b + E^a_b \bar{v}^b \bar{v}_a = 3j_2(x) + j_1(x) \,, \tag{46}$$

and are given by

$$3\bar{a}^6 H^2 = (\kappa T^a_b - C^a_b)\,\bar{v}_a\bar{v}^b$$
$$= \kappa\bar{\rho} - \frac{1}{2}\left[\mathcal{Q}^{(1)} + \mathcal{S}^{(1)}\right] , \tag{47a}$$

$$6\bar{a}^6\left(\frac{\ddot{\bar{a}}}{\bar{a}} + 3H^2\right) = (-\kappa T^a_b + C^a_b)\left(\bar{v}_a\bar{v}^b + \pi^b_a\right)$$
$$= -\kappa\left(\bar{\rho} + 3\bar{p}\right) + 2\left[\mathcal{Q}^{(1)} + \mathcal{Q}^{(2)} + \mathcal{S}^{(2)}\right] . \tag{47b}$$

Here Eqn. (47a) is the modified Friedmann equation and Eqn. (47b) the modified Raychaudhuri equation (in the vol0ume preserving gauge on $\bar{\mathcal{M}}$). We have used Eqn. (27), with the overbar on ρ and p reminding us that they are expressed in terms of the nonsynchronous time $\bar{t}$, and we have defined the correlation terms

$$\mathcal{Q}^{(1)} = \bar{a}^6\left[\frac{2}{3}\left(\langle\frac{1}{h}\Theta^2\rangle - \frac{1}{\bar{a}^6}({}^{\mathrm{F}}\Theta^2)\right) - 2\langle\frac{1}{h}\sigma^2\rangle\right] ;$$
$$\frac{1}{\bar{a}^6}({}^{\mathrm{F}}\Theta^2) = (3H)^2 , \tag{48a}$$

$$\mathcal{S}^{(1)} = \frac{1}{\bar{a}^2}\delta^{AB}\left[\langle\,{}^{(3)}\Gamma^J_{AC}\,{}^{(3)}\Gamma^C_{BJ}\rangle\right.$$
$$\left. -\langle\partial_A(\ln\sqrt{h})\partial_B(\ln\sqrt{h})\rangle\right] , \tag{48b}$$

$$\mathcal{Q}^{(2)} = \bar{a}^6\langle\frac{1}{h}\Theta^A_B\Theta^B_A\rangle - \frac{1}{\bar{a}^2}\delta^{AB}\langle\Theta_{AJ}\Theta^J_B\rangle, \tag{48c}$$

$$\mathcal{S}^{(2)} = \bar{a}^6\langle\frac{1}{h}h^{AB}\partial_A(\ln\sqrt{h})\partial_B(\ln\sqrt{h})\rangle$$
$$-\frac{1}{\bar{a}^2}\delta^{AB}\langle\partial_A(\ln\sqrt{h})\partial_B(\ln\sqrt{h})\rangle . \tag{48d}$$

In defining $\mathcal{Q}^{(1)}$ we have used the relation $\Theta^2 - \Theta^A_B\Theta^B_A = (2/3)\Theta^2 - 2\sigma^2$. $\mathcal{Q}^{(1)}$ and $\mathcal{Q}^{(2)}$ are correlations of the extrinsic curvature, whereas $\mathcal{S}^{(1)}$ and $\mathcal{S}^{(2)}$ are correlations restricted to the intrinsic 3-geometry of the spatial slices of $\mathcal{M}$. Since the components of C^a_b are not explicitly constrained we can treat the combinations $(1/2)(\mathcal{Q}^{(1)} + \mathcal{S}^{(1)}) = -C^0_0$ and $2(\mathcal{Q}^{(1)} + \mathcal{Q}^{(2)} + \mathcal{S}^{(2)}) = (C^A_A - C^0_0)$ as independent, subject only to the differential constraints which we will come to below.

As discussed in the beginning of Section 5, the remaining components of C^a_b must be set to zero, giving constraints on the underlying inhomogeneous geometry. In coordinate independent language, these constraints read

$$\pi^b_k C^a_b\bar{v}_a = 0 = \pi^k_a C^a_b\bar{v}^b ;$$
$$\pi^i_a\pi^b_k C^a_b - \frac{1}{3}\pi^i_k\left(\pi^b_a C^a_b\right) = 0 . \tag{49}$$

Eqns. (49) reduce to the following for our specific choice of volume preserving coordinates,

$$C^0_A = 0 \;\;;\;\; C^A_0 = 0 \;\;;\;\; C^A_B - \frac{1}{3}\delta^A_B(C^J_J) = 0\,, \tag{50}$$

It can be shown that the VPC assumption $N = h^{-1/2}$ reduces the correlations $\mathcal{Q}^{(2)}$ and $\mathcal{S}^{(2)}$ defined in Eqns. (48c) and (48d), as well as several terms in the explicit expansion of Eqn. (50), to the form

$$\frac{1}{\langle g_{00}\rangle}\langle g_{00} g^{AB}\Gamma^{a_1}_{b_1c_1}\Gamma^{i_1}_{j_1k_1}\rangle - \langle g^{AB}\rangle\langle\Gamma^{a_2}_{b_2c_2}\Gamma^{i_2}_{j_2k_2}\rangle\,. \tag{51}$$

It can be shown that

$$\begin{aligned}\langle g_{00} g^{AB}\Gamma^a_{bc}\Gamma^i_{jk}\rangle &= \langle g_{00} g^{AB}\rangle\langle\Gamma^a_{bc}\Gamma^i_{jk}\rangle \\ &= -\langle\frac{h^{AB}}{h}\rangle\langle\Gamma^a_{bc}\Gamma^i_{jk}\rangle\,.\end{aligned} \tag{52}$$

An interesting point is that the VPC assumption $N = h^{-1/2}$ further allows us to assume $\langle h^{AB}/h\rangle = \langle h^{AB}\rangle\langle 1/h\rangle$ consistently with the formalism. Using Eqn. (35) this gives us

$$\langle\frac{h^{AB}}{h}\rangle = \frac{1}{\bar{a}^6}\langle h^{AB}\rangle\,. \tag{53}$$

This shows that the correlation terms $\mathcal{Q}^{(2)}$ and $\mathcal{S}^{(2)}$ in fact vanish,

$$\mathcal{Q}^{(2)} = 0 = \mathcal{S}^{(2)}\,, \tag{54}$$

and leads to some remarkable cancellations in Eqns. (50), which simplify to give

$$\delta^{JK}\left[\langle\sqrt{h}\Theta_{JB}\,{}^{(3)}\Gamma^B_{AK}\rangle - -\langle\sqrt{h}\Theta_{JK}\,{}^{(3)}\Gamma^B_{AB}\rangle\right] = 0\,, \tag{55a}$$

$$\delta^{JK}\langle\frac{1}{\sqrt{h}}\Theta^B_K\,{}^{(3)}\Gamma^A_{JB}\rangle - \delta^{AJ}\langle\frac{1}{\sqrt{h}}\Theta^K_K\,{}^{(3)}\Gamma^B_{JB}\rangle = 0\,, \tag{55b}$$

$$\begin{aligned}\delta^{JK}\langle\,{}^{(3)}\Gamma^A_{JC}\,{}^{(3)}\Gamma^C_{KB}\rangle - \delta^{AJ}\langle\,{}^{(3)}\Gamma^C_{JC}\,{}^{(3)}\Gamma^K_{BK}\rangle \\ = \frac{1}{3}\delta^A_B\left(\bar{a}^2\mathcal{S}^{(1)}\right)\,.\end{aligned} \tag{55c}$$

These simplifications are solely a consequence of assuming that the inhomogeneous metric in the volume preserving gauge averages out to give the FLRW metric in standard form. In general, these simplifications will not occur when the standard FLRW metric arises from an arbitrary choice of gauge for the inhomogeneous metric.

In order to come as close as possible to the standard approach in Cosmology, we will now rewrite the scalar equations (47) (which are the cosmologically relevant ones) after performing the transformation given in Eqn. (39) in order to get the FLRW metric to the form

$${}^{(\bar{\mathcal{M}})}ds^2 = -d\tau^2 + a^2(\tau)\delta_{AB}dy^A dy^B \;\;;\;\; a(\tau) = \bar{a}(\bar{t}(\tau))\,. \tag{56}$$

Since Eqns. (47) are *scalar* equations, this transformation only has the effect of reexpressing all the terms as functions of the synchronous time τ. Although the transformation will change the explicit form of the coordination bivector $\mathcal{W}^{a'}_j$, this change involves only the time coordinate, and

in the spatial averaging limit there is no difference between averages computed in the VPCs and those computed after the time redefinition. This again emphasizes the importance of the spatial averaging limit of spacetime averaging, if we are to succeed operationally in explicitly displaying the correlations as corrections to the standard cosmological equations. The correlation terms in Eqns. (48) are therefore still interpreted with respect to the volume preserving gauge, but are treated as functions of τ. For the scale factor on the other hand, we have

$$\bar{a}^3 H = \frac{1}{a}\frac{da}{d\tau} \equiv H_{\mathrm{FLRW}} \;\; ; \;\; \bar{a}^6\left(\frac{\ddot{\bar{a}}}{\bar{a}} + 3H^2\right) = \frac{1}{a}\frac{d^2a}{d\tau^2}\,. \tag{57}$$

Further writing

$$\rho(\tau) = \bar{\rho}(\bar{t}(\tau)) \;\; ; \;\; p(\tau) = \bar{p}(\bar{t}(\tau))\,, \tag{58}$$

equations (47) become

$$H^2_{\mathrm{FLRW}} = \frac{8\pi G_N}{3}\rho - \frac{1}{6}\left[\mathcal{Q}^{(1)} + \mathcal{S}^{(1)}\right], \tag{59a}$$

$$\frac{1}{a}\frac{d^2a}{d\tau^2} = -\frac{4\pi G_N}{3}\left(\rho + 3p\right) + \frac{1}{3}\mathcal{Q}^{(1)}\,. \tag{59b}$$

We emphasize that the quantities $\mathcal{Q}^{(1)}$ and $\mathcal{S}^{(1)}$, defined in Eqns. (48a) and (48b) as correlations in the *volume preserving* gauge, are to be thought of as functions of the *synchronous* time τ, where the coordinate τ itself was defined *after* the spatial averaging. Such an identification is justified since we are dealing with scalar combinations of these quantities. Note that $\mathcal{Q}^{(1)}$ and $\mathcal{S}^{(1)}$ can be treated independently, apart from the constraints imposed by conservation conditions, which we turn to next. These conservation conditions can be decomposed into a scalar part and a 3-vector part, given respectively by

$$\bar{a} - rv^b C^a_{b;a} = 0 \;\; ; \;\; \pi^b_k C^a_{b;a} = 0\,. \tag{60}$$

In the synchronous gauge (56) for the FLRW metric, the scalar equation reads

$$\left(\partial_\tau \mathcal{Q}^{(1)} + 6H_{\mathrm{FLRW}}\mathcal{Q}^{(1)}\right) + \left(\partial_\tau \mathcal{S}^{(1)} + 2H_{\mathrm{FLRW}}\mathcal{S}^{(1)}\right) = 0\,. \tag{61}$$

We recall that this equation is a consequence of setting the correlation 3-form and the correlation 4-form to zero, and it relates the evolution of $\mathcal{Q}^{(1)}$ and $\mathcal{S}^{(1)}$. The 3-vector equation (on imposing the first set of conditions in Eqn. (49)) simply gives $\partial_\tau C^\tau_A = 0$, so that $C^\tau_A = 0 = \text{constant}$, which also implies that $C^A_\tau = 0 = \text{constant}$ and hence this equation gives nothing new. (We have used the relations $C^0_0 = C^\tau_\tau$, $C^0_A = \bar{a}^3 C^\tau_A$ and $C^A_0 = (1/\bar{a}^3)C^A_\tau$ where 0 denotes the nonsynchronous time coordinate $\bar{t}$.)

The cosmological equations (59), along with the constraint equations (55) and (61) are the key results of this section. Subject to the acceptance of the volume preserving gauge on the underlying manifold $\mathcal{M}$ they can in principle be used to study the role of the correction terms resulting from spatial averaging.

6.1 A comparison with the averaging formalism of Buchert

The averaging formalism developed by Buchert is based exclusively on the manifold $\mathcal{M}$, and there is no analog of the averaged manifold $\bar{\mathcal{M}}$ in this scheme. Given an inhomogeneous metric on $\mathcal{M}$ one

takes the trace of the Einstein equations in the *inhomogeneous* geometry, and carries out a spatial averaging of the inhomogeneous scalar equations.

For ease of comparison, we again recall in brief Buchert's construction, by first writing down the averaged equations for the simplest case of pressureless and irrotational inhomogeneous dust. The metric can be written in synchronous and comoving gauge as

$$ds^2 = -dt^2 + b_{AB}(\mathbf{x}, t)dx^A dx^B \,. \tag{62}$$

The Einstein equations can be split into a set of scalar equations and a set of vector and traceless tensor equations. The scalar equations are the Hamiltonian constraint (63a) and the evolution equation for Θ (63b),

$$\mathcal{R} + \frac{2}{3}\Theta^2 - 2\sigma^2 = 16\pi G\rho \,, \tag{63a}$$

$$\mathcal{R} + \partial_t\Theta + \Theta^2 = 12\pi G\rho \,, \tag{63b}$$

where $\mathcal{R}$ is the Ricci scalar of the 3-dimensional hypersurface of constant t, Θ and σ^2 are the expansion scalar and the shear scalar defined earlier and ρ is the inhomogeneous matter density of the dust. Note that all quantities in Eqns. (63) generically depend on both position $\mathbf{x}$ and time t. Eqns. (63a) and (63b) can be combined to give Raychaudhuri's equation

$$\partial_t\Theta + \frac{1}{3}\Theta^2 + 2\sigma^2 + 4\pi G\rho = 0 \,. \tag{64}$$

The continuity equation $\partial_t\rho = -\Theta\rho$ which gives the evolution of ρ, is consistent with Eqns. (63a), (63b). Only scalar Einstein equations are considered, since the spatial average of a scalar quantity can be defined in a gauge covariant manner, within a given foliation of space-time. We return to this point below. For the space-time described by (62), the spatial average of a scalar $\Psi(\mathbf{x}, t)$ over a *comoving* domain $\mathcal{D}$ at time t is defined by

$$\langle\Psi\rangle_{\mathcal{D}} = \frac{1}{V_{\mathcal{D}}}\int_{\mathcal{D}} d^3x\sqrt{b}\,\Psi \,, \tag{65}$$

where b is the determinant of the 3-metric b_{AB} and $V_{\mathcal{D}}$ is the volume of the comoving domain given by $V_{\mathcal{D}} = \int_{\mathcal{D}} d^3x\sqrt{b}$. Spatial averaging is, by definition, not generally covariant. Thus the choice of foliation is relevant, and should be motivated on physical grounds. In the context of cosmology, averaging over freely-falling observers is a natural choice, especially when one intends to compare the results with standard FLRW cosmology. Following the definition (65) the following commutation relation then holds

$$\partial_t\langle\Psi\rangle_{\mathcal{D}} - \langle\partial_t\Psi\rangle_{\mathcal{D}} = \langle\Psi\Theta\rangle_{\mathcal{D}} - \langle\Psi\rangle_{\mathcal{D}}\langle\Theta\rangle_{\mathcal{D}} \,, \tag{66}$$

which yields for the expansion scalar Θ

$$\partial_t\langle\Theta\rangle_{\mathcal{D}} - \langle\partial_t\Theta\rangle_{\mathcal{D}} = \langle\Theta^2\rangle_{\mathcal{D}} - \langle\Theta\rangle^2_{\mathcal{D}} \,. \tag{67}$$

Introducing the dimensionless scale factor $a_{\mathcal{D}} \equiv (V_{\mathcal{D}}/V_{\mathcal{D}i})^{1/3}$ normalized by the volume of the domain $\mathcal{D}$ at some initial time t_i, we can average the scalar Einstein equations (63a), (63b) and the continuity equation to obtain

$$\partial_t\langle\rho\rangle_{\mathcal{D}} = -\langle\Theta\rangle_{\mathcal{D}}\langle\rho\rangle_{\mathcal{D}} \quad ; \quad \langle\Theta\rangle_{\mathcal{D}} = 3\frac{\partial_t a_{\mathcal{D}}}{a_{\mathcal{D}}} \,, \tag{68a}$$

$$\left(\frac{\partial_t a_{\mathcal{D}}}{a_{\mathcal{D}}}\right)^2 = \frac{8\pi G}{3}\left\langle\rho\right\rangle_{\mathcal{D}} - \frac{1}{6}\left(\mathcal{Q}_{\mathcal{D}} + \left\langle\mathcal{R}\right\rangle_{\mathcal{D}}\right)\,, \tag{68b}$$

$$\left(\frac{\partial_t^2 a_{\mathcal{D}}}{a_{\mathcal{D}}}\right) = -\frac{4\pi G}{3}\left\langle\rho\right\rangle_{\mathcal{D}} + \frac{1}{3}\mathcal{Q}_{\mathcal{D}}\,. \tag{68c}$$

Here, the 'kinematical backreaction' $\mathcal{Q}_{\mathcal{D}}$ is given by

$$\mathcal{Q}_{\mathcal{D}} \equiv \frac{2}{3}\left(\left\langle\Theta^2\right\rangle_{\mathcal{D}} - \left\langle\Theta\right\rangle_{\mathcal{D}}^2\right) - 2\left\langle\sigma^2\right\rangle_{\mathcal{D}} \tag{69}$$

and is a spatial constant over the domain $\mathcal{D}$.

A necessary condition for (68c) to integrate to (68b) takes the form of the following differential equation involving $\mathcal{Q}_{\mathcal{D}}$ and $\left\langle\mathcal{R}\right\rangle_{\mathcal{D}}$,

$$\partial_t\mathcal{Q}_{\mathcal{D}} + 6\frac{\partial_t a_{\mathcal{D}}}{a_{\mathcal{D}}}\mathcal{Q}_{\mathcal{D}} + \partial_t\left\langle\mathcal{R}\right\rangle_{\mathcal{D}} + 2\frac{\partial_t a_{\mathcal{D}}}{a_{\mathcal{D}}}\left\langle\mathcal{R}\right\rangle_{\mathcal{D}} = 0\,. \tag{70}$$

The equations above describe the essence of Buchert's averaging formalism, for the dust case. We note that the remaining eight Einstein equations for the inhomogeneous geometry, which are not scalar equations, are not averaged. These are the five evolution equations for the trace-free part of the shear,

$$\partial_t\left(\sigma^A_B\right) = -\Theta\sigma^A_B - \mathcal{R}^A_B + \frac{2}{3}\delta^A_B\left(\sigma^2 - \frac{1}{3}\Theta^2 + 8\pi G\rho\right)\,. \tag{71}$$

and the three equations relating the spatial variation of the shear and the expansion,

$$\sigma^A_{B||A} = \frac{2}{3}\Theta_{||B}\,. \tag{72}$$

Here, $\mathcal{R}^A_B$ is the spatial Ricci tensor and, in Buchert's notation, a $||$ denotes covariant derivative with respect to the 3-metric.

In analogy with the dust case, Buchert's averaging formalism can be applied to the case of a perfect fluid, by starting from the metric

$$ds^2 = -N^2dt^2 + b_{AB}dx^Adx^B\,. \tag{73}$$

The averaged scalar Einstein equations for the scale factor $a_{\mathcal{D}}$ are

$$3\frac{\partial_t^2 a_{\mathcal{D}}}{a_{\mathcal{D}}} + 4\pi G\left\langle N^2\left(\rho + 3p\right)\right\rangle_{\mathcal{D}} = \bar{\mathcal{Q}}_{\mathcal{D}} + \bar{\mathcal{P}}_{\mathcal{D}}\,, \tag{74}$$

$$6H^2_{\mathcal{D}} - 16\pi G\left\langle N^2\rho\right\rangle_{\mathcal{D}} = -\bar{\mathcal{Q}}_{\mathcal{D}} - \left\langle N^2\mathcal{R}\right\rangle_{\mathcal{D}} \quad ; \quad H_{\mathcal{D}} = \frac{\partial_t a_{\mathcal{D}}}{a_{\mathcal{D}}}\,, \tag{75}$$

where the kinematical backreaction $\bar{\mathcal{Q}}_{\mathcal{D}}$ is given by

$$\bar{\mathcal{Q}}_{\mathcal{D}} = \frac{2}{3}\left(\left\langle\left(N\Theta\right)^2\right\rangle_{\mathcal{D}} - \left\langle N\Theta\right\rangle^2_{\mathcal{D}}\right) - 2\left\langle N^2\sigma^2\right\rangle_{\mathcal{D}}\,, \tag{76}$$

and the dynamical backreaction $\bar{\mathcal{P}}_{\mathcal{D}}$ is given by

$$\bar{\mathcal{P}}_{\mathcal{D}} = \left\langle N^2\mathcal{A}\right\rangle_{\mathcal{D}} + \left\langle\Theta\partial_t N\right\rangle_{\mathcal{D}}\,, \tag{77}$$

where $\mathcal{A} = \nabla_j(u^i\nabla_i u^j)$ is the 4-divergence of the 4-acceleration of the fluid. Eqn. (75) follows as an 0integral from Eqn. (74) if and only if the relation

$$\partial_t\mathcal{Q}_{\mathcal{D}} + 6H_{\mathcal{D}}\mathcal{Q}_{\mathcal{D}} + \partial_t\left\langle N^2\mathcal{R}\right\rangle_{\mathcal{D}} + 2H_{\mathcal{D}}\left\langle N^2\mathcal{R}\right\rangle_{\mathcal{D}} + 4H_{\mathcal{D}}\bar{\mathcal{P}}_{\mathcal{D}} - 16\pi G\left[\partial_t\left\langle N^2\rho\right\rangle_{\mathcal{D}} + 3H_{\mathcal{D}}\left\langle N^2\left(\rho+p\right)\right\rangle_{\mathcal{D}}\right] = 0\,, \tag{78}$$

is satisfied. There are also the unaveraged equations (which we do not display here) for the shear, analogous to the shear equations (71) and (72) for dust.

Buchert's approach is the only other approach, apart from Zalaletdinov's MG, which is capable of treating inhomogeneities in a nonperturbative manner, although it is limited to using only scalar quantities within a chosen $3+1$ splitting of spacetime. Buchert takes the trace of the Einstein equations in the *inhomogeneous* geometry, and averages these inhomogeneous scalar equations. In the context of Zalaletdinov's MG however, we have used the existence of the vector field $\bar{v}^a$ in the FLRW spacetime to construct scalar equations *after* averaging the full Einstein equations. As far as observations are concerned, it has been noted by Buchert and Carfora that the spatially averaged matter density $\langle\rho\rangle_{\mathcal{D}}$ defined by Buchert is *not* the appropriate observationally relevant quantity – the "observed" matter density (and pressure) is actually defined in a *homogeneous* space. Since we have done precisely this in Eqn. (27), we are directly dealing with the appropriate observationally relevant quantity in the MG framework.

Another important difference between the two approaches is the averaging operation itself. Buchert's spatial average, defined for scalar quantities, is given (for some scalar $\Psi(t,x^A)$) by (65) above. On the other hand the averaging operation we have been using (given by Eqn. (33) using the volume preserving gauge) is a limit of a spacetime averaging defined using the coordination bivector $\mathcal{W}_j^{a'}$, and is different from the one in Eqn. (65).

Most importantly though, Buchert's averaging scheme by itself does not incorporate the concept of an averaged manifold $\bar{\mathcal{M}}$ (although the work of Buchert and Carfora [2] does deal with 3-spaces of constant curvature). In a recent paper we had argued that Buchert's "effective scale factor" $a_{\mathcal{D}}(t) \equiv (V_{\mathcal{D}}(t)/V_{\mathcal{D}}(t_{in}))^{1/3}$ must be the scale factor for the metric of the averaged manifold, upto some corrections arising due to such effects as calculated by Buchert and Carfora. In the present work however, it is clear that such a suggestion is necessarily incomplete due to the presence of Eqns. (55) constraining the underlying geometry. These constraints are in general nontrivial and hence indicate that it is not sufficient to assume that the metric of the inhomogeneous manifold averages out to the FLRW form – there are additional conditions which the correlations must satisfy.

To our understanding, Buchert's averaging formalism is a valid aproach, even though it is based on a spatial averaging. A central difference from the MG approach is the issue of closure : not all the Einstein equations have been averaged in Buchert's approach, but only the scalar ones. This puts a constraint on the allowed solutions considered for the averaged equations: (68) for the dust case, and (74) and (75) for the fluid case. Solutions to these equations must necessarily be checked for consistency with the unaveraged equations for the shear. Further, averaging over successively larger scales can bring in additional corrections to the averaged equations, as discussed by Buchert and Carfora. Also, if one does not wish to identify Buchert's $a_{\mathcal{D}}$ with the scale factor in FLRW cosmology, one is compelled to develop a whole new set of ideas in order to try and compare theory with observation. On the other hand, if one does identify $a_{\mathcal{D}}$ with the scale factor, comparison with standard cosmology becomes more convenient, but this brings in additional constraints on the underlying inhomogeneous geometry. Thus our conclusion is that the Buchert formalism is a correct and tractable averaging scheme, provided all the caveats pointed out in this paragraph are

taken care of. Also, when these caveats have been taken care of correctly, the Buchert formalism is expected to give the same physical results as the MG approach. We recall that in the covariant MG approach also, once a spacetime geometry has been identified for the averaged manifold $\bar{\mathcal{M}}$, a gauge must be selected for the geometry on the underlying manifold, in order to explicitly compute the correction scalars for comparison with observation.

The advantage of the MG approach is that it accomplishes in a neat package what the Buchert approach, with its attendant caveats, sets out to do. In the MG approach, there are no unaveraged shear equations, because the trace of the Einstein equations has been taken after performing the averaging on the underlying geometry. Since the averaged geometry is FLRW, the shear is zero by definition. There is a natural metric on the averaged manifold by construction, the FLRW metric. The correlations satisfy additional constraints, given by Eqns. (55). Thus, once a gauge has been chosen and if one can overcome the computational complexity of the averaging operation, the cosmological equations derived by us in the MG approach are complete and ready for application, without any further caveats.

In spite of these differences, our equations (59) and (61) for the volume preserving gauge are strikingly similar to Buchert's effective FLRW equations and their integrability condition in the dust case; and in the case of general N, the role of Buchert's dynamical backreaction $\bar{\mathcal{P}}_{\mathcal{D}}$ in Eqns. (74) and (78) is identical to that of our combination of $(\tilde{\mathcal{P}}^{(2)} + \tilde{\mathcal{S}}^{(2)})$. Concentrating on the volume preserving case, the structure of the correlation $\mathcal{Q}^{(1)}$ is identical to Buchert's kinematical backreaction $\mathcal{Q}_{\mathcal{D}}$ (or $\bar{\mathcal{Q}}_{\mathcal{D}}$ in the general case). The correlation $\mathcal{S}^{(1)}$ appears in place of the averaged 3-Ricci scalar $\langle \mathcal{R} \rangle_{\mathcal{D}}$ in Buchert's dust equations. This is not unreasonable since Buchert's $\langle \mathcal{R} \rangle_{\mathcal{D}}$ can be thought of as $\langle \mathcal{R} \rangle_{\mathcal{D}} = 6k_{\mathcal{D}}/a_{\mathcal{D}}^2 +$ corrections, where $6k_{\mathcal{D}}/a_{\mathcal{D}}^2$ represents the 3-Ricci scalar on the averaged manifold which in our case is zero, and hence $\mathcal{S}^{(1)}$ represents the corrections due to averaging. Further, these similarities are in spite of the fact that our correlations were defined assuming that a *volume preserving* gauge averages out to the FLRW 3-metric in standard form, whereas Buchert's averaging is most naturally adapted to beginning with a *synchronous* gauge. This remarkable feature, at least to our understanding, does not seem to have any deeper meaning – it simply seems to arise from the structure of the Einstein equations themselves, together with our assumption $\mathbf{D}_{\bar{\Omega}} \mathbf{Z}^{a}{}_{b}{}^{i}{}_{j} = 0$. In the absence of this latter condition, one would have to consider the correlation 3- and 4-forms mentioned earlier, and the structure of the correlation terms and their "conservation" equations would be far more complicated.

An entirely different outlook towards his approach has been emphasized to us by Buchert. According to Buchert, the absence of an averaged manifold $\bar{\mathcal{M}}$ is not to be thought of as a 'caveat', but as a feature deliberately retained 'on purpose'. The actual inhomogeneous Universe is regarded by Buchert as the only fundamental entity, and the introduction of an averaged Universe is in fact regarded as an unphysical and unnecessary approximation. As we mentioned earlier, this is probably the most important difference between MG and Buchert's approach. In the latter, contact with observations is to be made by constructing averaged quantities, such as the scalars defined earlier in this section, and by introducing the expansion factor $a_{\mathcal{D}}$. The assertion here is that the averaging of *geometry*, as discussed in MG or in the Renormalization Group approach of Buchert and Carfora [2] is not an indispensable step in comparing the inhomogeneous Universe with actual observations. The need for averaging of geometry is to be physically separated from simply looking at effective properties (such as the constructed scalars) which can be defined for any inhomogeneous metric. Averaging of geometry becomes relevant if (i) an observer insists on interpreting the data in a FLRW template model, so that (s)he needs a mapping from the actual inhomogeneous slice and its average properties to the corresponding properties in this template, or (ii) one desires a

mock metric, to sort of have a thermodynamic effective metric to approximate the real one. In this context it should perhaps also be mentioned that the importance of a thin time-slice approximation of spacetime averaging (as opposed to a strict spatial averaging) has been stressed also by Buchert.

7 Perturbation theory, structure formation, and backreaction

We have in hand the machinery to ask the following question : Is cosmological perturbation theory stable against growth of backreaction? The answer must be found iteratively. Assume a background with perturbations on it, calculate the back-reaction, feed it in the right hand of the modified Friedmann equations to find the new background, and so on :

$$a^{(0)} \to \phi^{(0)} \to C^{(0)} \to a^{(1)} \to \phi^{(1)} \to \dots \tag{79}$$

Let the perturbed FLRW metric be

$$ds^2 = a^2 \left[-(1+2\phi)d\eta^2 + 2\omega_A dx^A d\eta + ((1-2\psi)\gamma_{AB} + \chi_{AB})\, dx^A dx^B \right] . \tag{80}$$

We work with a VPC which has no residual degrees of freedom. Further, this VPC is constructed by starting from the conformal Newtonian gauge, and by making a steady coordinate transformation. This ensures that all averaged quantities are gauge invariant. We evaluate the correlation scalars for a given initial power spectrum - standard CDM.

For a constant nonevolving potential $\phi(\vec{x})$, and with a power spectrum

$$\frac{k^3 P_{\phi i}(k)}{2\pi^2} = A(k/H_0)^{n_s - 1} , \tag{81}$$

the back reaction is

$$\frac{\mathcal{S}^{(1)}}{H_0^2} \sim -\frac{1}{a^2}(10^{-4}) . \tag{82}$$

The smallness of backreaction holds also for the exact sCDM model thus demonstrating the stability of perturbation theory against the growth of back-reaction.

This analysis ignores contribution of scales that have become fully nonlinear in matter density at late times and it is important to ask if structure formation can significantly modify large sacle dynamics.

We studied backreaction in a toy model of spherical collapse, using the LTB solution. The initial density is chosen to be

$$\rho(t_i, r) = \rho_{bi} \begin{cases} (1+\delta_*), & r < r_* \\ (1-\delta_v), & r_* < r < r_v \\ 1, & r > r_v , \end{cases} \tag{83}$$

We match the initial velocity and coordinate scaling to the global background solution, by requiring

$$R(t_i, r) = a_i r , \tag{84}$$

$$\dot{R}(t_i, r) = a_i H_i r , \tag{85}$$

For the FLRW background we consider an Einstein-deSitter (EdS) solution with scale factor and Hubble parameter given by

$$a(t) = (t/t_0)^{2/3} \quad ; \quad t_0 = 2/(3H_0)\,, \tag{86}$$
$$H(t) \equiv \dot{a}/a = 2/(3t)\,, \tag{87}$$

with t_0 denoting the present epoch. a_i fixes the initial time as

$$t_i = 2/(3H_0)a_i^{3/2}\,. \tag{88}$$

We use $a_i = 10^{-3}$, so that the initial conditions are being set around the CMB last scattering epoch. The mass function $M(r)$ and curvature function $k(r)$ in this LTB solution are given by

$$GM(r) = \frac{1}{2}H_0^2 r^3 \begin{cases} 1+\delta_*, & 0 < r < r_* \\ 1+\delta_v\left((r_c/r)^3 - 1\right), & r_* < r < r_v \\ 1+(\delta_v/r^3)\left(r_c^3 - r_v^3\right), & r > r_v\,, \end{cases} \tag{89}$$

where we have defined a "critical" radius r_c by the equation

$$\left(\frac{r_c}{r_*}\right)^3 = 1 + \frac{\delta_*}{\delta_v}\,. \tag{90}$$

The significance of r_c is brought out by $k(r)$:

$$k(r) = \frac{H_0^2}{a_i} \begin{cases} \delta_*, & r < r_* \\ \delta_v\left((r_c/r)^3 - 1\right), & r_* < r < r_v \\ (\delta_v/r^3)\left(r_c^3 - r_v^3\right), & r > r_v\,. \end{cases} \tag{91}$$

Since $\delta_*, \delta_v > 0$, we have $r_c > r_*$ by definition. The following possibilities arise :

If $r_c > r_v$, then $k(r) > 0$ for all r, and every shell will ultimately collapse, including the "void" region $r_* < r < r_v$. If $r_c < r_v$, then $k(r) > 0$ for $r < r_c$ and changes sign at $r = r_c$. Hence, the region $r_* < r < r_c$ will collapse even though it is underdense, while the region $r > r_c$ will expand forever. If $r_c = r_v$, then the "void" exactly compensates for the overdensity, and the universe is exactly EdS for $r > r_v$.

Transforming to the perturbed FLRW form : We want a coordinate transformation $(t, r) \to (\tau, \tilde{r})$ such that the metric in the new coordinates is

$$ds^2 = -(1+2\phi)d\tau^2 + a^2(\tau)(1-2\psi)\left(d\tilde{r}^2 + \tilde{r}^2 d\Omega^2\right)\,, \tag{92}$$

with at least the conditions

$$|\,\phi\,| \ll 1 \quad ; \quad |\,\psi\,| \ll 1\,, \tag{93}$$

being satisfied. Since t is the proper time of each matter shell, the quantity $\partial_t \tilde{r}$ is simply the velocity of matter in the $(\tau, \tilde{r})$ frame (which is comoving with the Hubble flow) :

$$\tilde{v} \equiv \frac{\partial \tilde{r}}{\partial t}\,, \tag{94}$$

is the radial comoving peculiar velocity of the matter shells in the $(\tau, \tilde{r})$ frame. We showed that the required transformation exists, *provided matter peculiar velocities remain small*, which is consistent with what has een shown by other authors, and is true for the observed Universe.

In the cosmological equations derived from Macroscopic Gravity we already have in place the formalism for calculating the backreaction when the metric is of the perturbed FLRW form. From there it follows that the backreaction is very small, in the nonlinear structure formation regime, provided matter peculiar velocities are small. It can be argued that this result is independent of the assumption of spherical symmetry in the toy model. The situation could be very different though, if there are dominant nonlinear structures in today's Universe, comparable to the Hubble radius.

7.1 Perturbation theory around a background - the shortwave approximation

Green and Wald [22] have recently given an analysis of the growth of metric perturbations, assuming that the metric is always close to a given background, although matter perturbations can be arbitrarily large. No averaging of an underlying spacetime geometry is done, and it is assumed that there is a homogeneity length scale at around 100 Mpc, much smaller than the Hubble radius. It is shown that if the small-scale motions of matter inhomogeneities are non-relativistic, the deviations from the background metric are small, and well-described by Newtonian gravity. This result tallies with what has een found by others before, including us. It is further shown that subject to the matter satisfying weak energy condition, the effect of small scale inhomogeneities on large scale dynamics is to produce an effective trace-free stress energy tensor. One might ask if this traceless nature of the correction has to do with no averaging over finite volumes being carried out.

Thus the assumption of non-relativistic peculiar velocities along with the assumption of a homogeneity scale much smaller than the Hubble radius strongly suggest a negligible effect of small scale inhomogeneities on the average large-scale dynamics. The first of these two assumptions is well supported by observations. There is no observational evidence against the second assumption, but nor is it firmly established by observations. If this assumption is correct, either a small cosmological constant, or a modification of general relativity on large scales, is indicated by the observed cosmic acceleration. If this assumption turns out to be not correct, the effect of inhomogeneities could be significant, and remains an important question for further investigation.

I would like to thank Aseem Paranjape for collaboration and extended discussions during the period 2006-2009 : our work on application of Macroscopic Gravity to cosmology would not have been possible without his ingenuity in simplifying the original system of equations. I would like to thank Friedrich Hehl for suggesting in the first place that we apply MG to cosmology. Correspondence and interactions with Roustam Zalaletdinov are gratefully acknowledged. I am also thankful to Thomas Buchert for correspondence in the early stages of this work. It is a pleasure to thank the organizers of the conference for their kind hospitality, and the conference participants for stimulating discussions.

The list of references below is far from exhaustive, and references to a large number of the original papers on the subject can be found in the review articles cited here.

References

[1] G. F. R. Ellis (2011) *Inhomogeneity effects in cosmology* arXiv:1103.2335 [astro-ph.CO]

[2] T. Buchert (2011) *Towards physical cosmology : focus on inhomogeneous geometry and its non-perturbative effects* arXiv:1103.2016 [gr-qc]

[3] A. Coley, J. Brannlund and J. Latta (2011) *Unimodular gravity and averaging* arXiv:1102.3456 [gr-qc]

[4] A. Ishibashi and R. M. Wald, *Can the acceleration of our universe be explained by the effects of inhomogeneities?*, Class. Quantum Grav. 23, 235 (2006)

[5] R. Zalaletdinov (1993) *Towards a macroscopic theory of gravity* Gen. Rel. Grav. 25, 673

[6] E. W. Kolb, S. Matarrese and A. Riotto (2006) *On cosmic acceleration without dark energy* New J. Phys. 8, 322

[7] V. Marra and A. Notari *Observational constraints on inhomogeneous cosmological models without dark energy* arXiv:1102.1015 [astro-ph.CO]

[8] S. Rasanen (2011) *Back-eaction : directions of progress* arXiv:1102.0405

[9] R. A. Sussman (2011) *A comprehensive study of back-reaction and effective acceleration in generic LTB dust models* arXiv:1102.2663 [gr-qc]

[10] D. Wiltshire (2011) *Gravitational energy as dark energy : cosmic stucture and apparent acceleration* arXiv:1102.2045

[11] Chris Clarkson and Roy Maartens (2010) *Inhomogeneity and the foundations of concordance cosmology* Class. Quantum Grav. 27, 124008

[12] Roy Maartens (2011) *Is the universe homogeneous* arXiv:1104.1300 [astro-ph.CO]

[13] P. J. E. Peebles (2009) *Phenomenology of the invisible universe* arXiv:0910.5142

[14] Aseem Paranjape and T. P. Singh (2007) *The spatial averaging limit of covariant macroscopic gravity : scalar corrections to the cosmological equations* Phys. Rev. D76:044006

[15] Aseem Paranjape and T. P. Singh (2008) *Structure formation, back-reaction and weak gravitational fields* JCAP 0803:023

[16] Aseem Paranjape (2008) *Backreaction of cosmological perturbations in covariant macroscopic gravity* Phys. Rev. D78:063522

[17] Aseem Paranjape and T. P. Singh (2008) *Cosmic inhomogeneities and the average cosmological dynamics* Phys. Rev. Lett. 101:181101

[18] R. Zalaletdinov (1992) *Averaging out the Einstein equations and macroscopic spacetime geometry* Gen. Rel. Grav. 24, 1015

[19] R. Zalaletdinov (1997) *Averaging problem in general relativity, macroscopic gravity, and using Einstein's equations in cosmology* Bull. Astron. Soc. India 25, 401

[20] Marc Mars and R. Zalaletdinov (1997) *Spacetime averages in macroscopic gravity and volume preserving coordinates* J. Math. Phys. 38, 4741

[21] A. A. Coley, N. Pelavas and R. Zalaletdinov (2005) *Cosmological solutions in macroscopic gravity* Phys. Rev. Lett. 95:151102

[22] S. R. Green and R. M. Wald (2010) *A new framework for analyzing the effects of small scale inhomogeneities in cosmology* arXiv:1011.4920 [gr-qc]

[23] Aseem Paranjape (2009) *The averaging problem in cosmology* Ph. D. Thesis, arXiv:0906.3165

[24] E. R. Siegel and J. N. Fry *The effect of inhomogeneities on cosmic expansion* Ap. J. 628 : L1-L4 (2005).

PART III

THE APPARENT ACCELERATION

OF THE EXPANSION OF THE UNIVERSE

Cosmic Acceleration: what do data actually tell us?

BLANCHARD Alain
IRAP, 14 Av. E.Belin, 31400 Toulouse, FRANCE

April 1, 2011

Abstract

Evidence for an accelerated expansion of the universe as it has been revealed ten years ago by the Hubble diagram of distant type Ia supernovae represents the lattest revolution of modern cosmology with profound impact for fundamental physics. The construction of a scientfic model of the universe is probably one of the most fascinating success of XXth century science. During its construction, there has been regular debates regarding whether the whole construction being scientific. Indeed, existing evidence for the big bang picture, including its modern version the Λ CDM picture, comes from astrophysical observations. It is therefore interesting and essantial to critically examine the present situation of the astrophysical observations and the possible limitation in their interpretation. In this paper, the main various observational probes at the fundation of the standard view are presented as well as the standard framework to interpret them with special attention to the complex astrophysics and theoretical hypotheses that may limit robust interpretation. It is concluded that, even when scrutinized with sceptical eyes, the evidence for a homogenous accelerated universe, governed by standard Friedman-Lemaître equations, is robust. Therefore the standard Λ CDM picture has to be regarded as the most successful scientific representation of the universe by now, possibly being the only one clearly consistent with the whole family of observations relevant to cosmology. The fact that this model could in principle be easily falsified makes it a very good scientific theory. The history of the cosmological constant is cosmology would deserve a long discussion, but in order to make a long story short, we can summarize it by saying that the discovery of the expansion of the universe has convinced most of researchers in this field until 1998, including Einstein himself, to consider the addition of this term unnecessary, while nowadays, no alternative to the accelerated Universe had achieved the same amount of scientific successes. Understanding the origin of acceleration is probably one of the most challenging problem of fundamental physics.

1 Introduction

Cosmology as a science has grown after the discovery of the cosmic background radiation, less than fifty years ago. Since that time impressive amount of progresses was achieved, leading to remarkable consequences for fundamental physics, including constraints on physics at energy well above to what is reacheable in accelerators, the evidence for dark matter and recently the evidence for dark energy. The general picture, the "Big Bang", including the inflation Λ CDM model is now recognized as the successful scientific representation of the world at the large scales (in space and in time) we can measure.

2 Basics of Friedmann-Lemaître Models

The fundamental idea of the geometrical theory of gravity starts from the fact that we can assign four coordinates to any event observed in our vicinity, for instance in Cartesian coordinates : (x, y, z, t). Locally, space appears to be flat, or nearly so. However this does not prejudge of the geometry of space at larger scales : local observations put us in the same situation that led people to think the earth was flat : the fact that we can describe our vicinity by a flat map does not prejudge of the actual geometry on larger scales. Let us take the line element of a homogeneous 3D space which can be shown[1] to be :

$$dl^2 = \tilde{r}^2(d\theta^2 + \sin^2\theta d\phi^2) + \frac{d\tilde{r}^2}{1 - k\left(\frac{\tilde{r}}{R}\right)^2} \tag{1}$$

where k is $-1, 0, 1$ according to whether space is hyperbolic, flat or spherical. R is a characteristic size (in the spherical case, that is the radius of the 3D-sphere embedded in a 4D space).

We then add the time as the fourth coordinate to build the equivalent of the Minkowski space-time element of special relativity and get the Robertson-Walker (RW) line element after the change of variables $\frac{\tilde{r}}{R} \to r$:

$$ds^2 = -c^2dt^2 + R(t)^2[r^2(d\theta^2 + \sin^2\theta d\phi^2) + \frac{dr^2}{1 - kr^2}] \tag{2}$$

2.1 Homogeneity

The starting hypothesis of modern standard cosmological model is the asumption that the universe is homegenous on large scale. This is the fundamental hypothesis, which I personnally would trace back to Giordano Bruno, even if not phrased in modern words. This is now called the Einstein cosmologcial principle and justifies the use the RW metric to describe our universe. It is of course vital to check this asumption from observations. There is a long tradition to contest this asumption. For instance Carl Charlier proposed a "fractal" picture at the beginning of XXth century. More recently, G.F.R. Ellis has since several years developped a point of view that observations do not prove that the universe is actually homognenous on large scale and therefore the starting hypothesis of the standard picture might be wrong. Therefore it might be that the actual universe is inhomogenous and that we describe it incorrectly by a homogenous picture. All this is right but is the very common situation in physics: any theory is based on a number of asumptions which cannot be directly proven by the observations to be right. In fact the purpose of physics is not to tell what is right or not... What is wanted in physics is to have a description which is simple and making as many predictions as possible that can be tested. In this respect the RW metric has been successive and there is almost no competitor on the market!

A further misconception about this question lies in the way it is phrased in order to use the RW metric. The homogneity condition is generally formulated on the matter fluctuations, i.e. that matter fluctuations become small on large scale :

$$\lim_{R \to +\infty} \delta\rho_x(R) = \overline{\rho} \tag{3}$$

[1] It is an instructive exercise to start from an Euclidean 4D space x, y, z, u and derive the line elements dl^2 on the 3D sphere $(x^2 + y^2 + z^2 + u^2 = R^2)$ in internal spherical coordinates $(\tilde{r} = \sqrt{x^2 + y^2 + z^2}, \theta, \phi)$.

for any location x in the universe ($\overline{\rho}$ being independent of the location x). This is not correct. What is actually wanted is that the fluctuations in the *metric* δh associated to the density fluctuations are decreasing to zero on large scale:

$$\lim_{R\to+\infty} \delta h(R) = 0 \tag{4}$$

that is:

$$\lim_{R\to+\infty} G\delta\rho(R)R^2 = 0 \tag{5}$$

which is a condition much more demanding than homogeneity of the matter, whith little hope that observations will prove this condition to be satisfied, because even if one proves the homogeneity of matter distribution condition (Eq. 4), the second condition (Eq. 5) will obviously not automatically be satisfied. This means that we may very probably never prove directly that condition (5) is actually satisfied! However, it does not mean that the theory will not be satisfying: as long as the model is predictive and these predictions are verified by observations, the scientific attitude is to consider that the model meets no trouble. Of course alternative views should be encouraged and tested, but it is their duty to prove that they can offer an alternative as viable as the standard picture! Even in this case, their scientific merit does not reach that of the standard picture as long as they did not capitalize as many verification of their predictions. By present days, only inhomogeneous Lemaître-Tolman-Bondi models offer an alternative to the standard picture within general relativity, without having achieved the same amount of success in predictions/validations.

2.2 Topology

The Robertson-Walker line element describes the local shape of space : the curvature (i.e. the value of k/R^2) is only a local property of space, but does not tell us about the *global* shape of space. For instance, the Euclidean plane is an infinite flat surface while the surface of a cylinder is a 2D-space which is flat everywhere but is finite in one direction. GR in principle allows us to derive the local geometry of space and its dynamics, but does not specify the global topology of space. Only direct observations would allow to test what the topology actually is. Of course this will not be possible on scales much larger than what can be observed (the horizon). We can therefore hope to prove that the Universe is finite, if it is small enough, but we could not know whether we are in a finite Universe of which the scale is larger than the horizon, or whether we are in an infinite Universe. The interest in the topic of the cosmic topology, with possible observational signature, has been recently revived [14, 19]

2.3 Important quantities needed for observations

In this section we only need to work in the framework of a geometrical theory of space-time, in which the trajectories of light rays are assumed to be the null geodesics. Let us have a comoving spherical coordinate system (r, θ, ϕ, t) the observer being at the origin of the spatial coordinates $(r = 0, \theta = 0, \phi = 0, t_0)$, let assume that the observed source is emitting light at the coordinates $(r_{\rm S}, \theta = 0, \phi = 0, t_{\rm S})$, and let $r(t)$ be the trajectory of the emitted photons. As this trajectory is a null geodesic, we have:

$$c^2dt^2 - R^2(t)\frac{dr^2}{1-kr^2} = 0 \tag{6}$$

so the variables can be separated and the integration over r is analytical:

$$\int_{t_S}^{t_0} \frac{cdt}{R(t)} = \int_0^{r_S} \frac{dr}{(1-kr^2)^{1/2}} = S_k^{-1}(r_S) \tag{7}$$

with:

$$S_k(r_S) = \sin(r_S) \text{if} k = +1 r_S \text{if} k = 0 \sinh(r_S) \text{if} k = -1 \tag{8}$$

When the distance is small with respect to R_0 we just have $S_k^{-1}(r) \sim r$.

2.4 The Redshift

In order to derive the observed frequency ν_0 of the light from a source emitted at the frequency ν, we consider the trajectory of a second light ray emitted at the time $t_S + \frac{1}{\nu}$. As the source is comoving its coordinate is unchanged and we have:

$$S_k^{-1}(r_S) = \int_{t_S}^{t_0} \frac{cdt}{R(t)} = \int_{t_S+1/\nu}^{t_0+1/\nu_0} \frac{cdt}{R(t)} \tag{9}$$

which implies:

$$\frac{\nu_0}{\nu} = \frac{\lambda_S}{\lambda_0} = \frac{R_S}{R_0} = \frac{1}{1+z} \tag{10}$$

where z is the redshift. This is the standard formula for the cosmological shift of the frequencies. This result shows that the redshift z is a natural consequence of the expansion.

2.5 The proper distance

In GR, space changes with time, and there is no proper time, so that the "intuitive" notion of distance between two points is not a well defined quantity. Therefore the various methods to measure the distance between an observer and a given source give different answers. The proper distance – between the source and the observer – can be seen as a distance measured by a set of rulers at time t. The distance element is given by :

$$dl^2 = ds^2 = R(t)^2 \frac{dr^2}{1-kr^2} \tag{11}$$

so that the proper distance is :

$$D_p = R(t) S_k^{-1}(r_S) \tag{12}$$

The fact that this distance changes with time is the direct consequence of the expansion of the Universe. We can now examine how this length changes with time :

$$\dot{D}_p = \dot{R} S_k^{-1}(r) \tag{13}$$

so that the source is *actually receding* from the observer with a speed:

$$V = \dot{D}_p = \frac{\dot{R}}{R} D_p = H D_p \tag{14}$$

The fact that this speed could be larger than the speed of light should not be considered as a problem: this speed can be measured but cannot transport information faster than light. When the distance is small, the Doppler frequency shift is :

$$\frac{\delta\nu}{\nu} = \frac{\dot{R}}{R}\delta t = H\frac{D}{c} = \frac{V}{c} \tag{15}$$

so that the shift is the one corresponding to the Doppler shift associated with the above velocity. For large distances, the total shift results from the product of small Doppler shifts and the redshift is therefore purely kinematic. The physical nature of the expansion has been recently the subject of interesting discussions [1, 21, 8].

Comoving distances It is sometimes useful to refer to comoving distances[2]. The comoving distance $D^c(z)$ associated to the distance $D(z)$ is :

$$D^c(z) = \frac{R_0}{R}D(z) = (1+z)D(z) \tag{16}$$

In the case of the proper distance, this becomes:

$$D_p^c(z) = R_0 S_k^{-1}(r) = \int_{t_S}^{t_0} \frac{cdt}{R(t)/R_0} = c\int_0^z \frac{dz}{H(z)} \tag{17}$$

2.6 The angular distance

Let us suppose that we observe a ruler orthogonal to the line of sight. The extremities of the ruler have the coordinates $(r, 0, 0, t_S)$ and $(r, \theta, 0, t_S)$. The proper length l between the extremities is:

$$l^2 = ds^2 = R(t_S)^2 r^2 \theta^2 \tag{18}$$

which provides the relation between the angle θ and the length l and thereby the angular distance defined by:

$$D_{\rm ang} = \frac{l}{\theta} = R(t_S)r \tag{19}$$

2.7 The luminosity distance

Let us assume that we observe a source with an absolute luminosity L through a telescope with a diameter d and let us choose a coordinates system which is centered on the source. Let θ be the angle between two rays reaching two points diametrically opposite on the telescope. We have $d = R(t_0)r\theta$. The energy emitted by the source that reaches the telescope is :

$$s = \frac{L}{4\pi} \times \frac{\pi\theta^2}{4} \tag{20}$$

When observed, the energy of photons has been shifted by $1/(1+z)$ but also the frequency at which they arrive is reduced by the same factor. Therefore the flux (energy per unit time and unit surface) one gets is:

$$f = \frac{s}{\pi l^2/4}\frac{1}{(1+z)^2} = \frac{L}{4\pi R(t_0)^2 r^2 (1+z)^2} = \frac{L}{4\pi D_{\rm lum}^2} \tag{21}$$

[2]This could also be confusing!

This relation provides the luminosity distance:

$$D_{\rm lum} = R(t_0)r(1+z) = R(t_{\rm S})r(1+z)^2 = D_{\rm ang}(1+z)^2 \tag{22}$$

2.8 Dynamics

The function $R(t)$ which appears in the RW line element, is totally independent of any further geometrical consideration. It can be specified only within a theory of gravity. The basic equation of GR relates the geometrical tensor G_{ij} to the energy-momentum tensor T_{ij}

$$G_{ij} = R_{ij} - \frac{1}{2}g_{ij}R = 8\pi G T_{ij} \tag{23}$$

where g_{ij} is the metric tensor, R_{ij} is the Ricci tensor, R the Ricci scalar. For a perfect fluid, there exists a coordinates system, called the comoving coordinates, in which the matter is at rest and the tensor T_{ij} is diagonal with $T_{00} = \rho$ and $T_{11} = T_{22} = T_{33} = p$, ρ being the density and p the pressure. A fundamental aspect of GR is that the source of gravity includes explicitly a term coming from the pressure : $\rho + 3p/c^2$. Finally, there is an analog of the Gauss theorem, that is the Birkhoff's theorem [6][3]: if the matter distribution is spherical then the evolution of the radius of a given shell of matter depends only on its content.

From the above rules, we can easily derive the equation for $R(t)$. Let us consider a spherical region of radius a in a homogeneous distribution of matter. The equivalent Newtonian acceleration is :

$$\frac{d^2a}{dt^2} = g \tag{24}$$

with the acceleration being generated by the "mass" $M(a)$ of the above spherical region :

$$\ddot{a} = g = -\frac{GM(a)}{a^2} = -\frac{4}{3}\pi G(\rho + 3p/c^2)a \tag{25}$$

The density term includes the effect of kinetic energy ($E = mc^2$!). Writing total energy (E_t) conservation inside the volume of the sphere from elementary thermodynamics gives :

$$d(E_t) = d(\rho V c^2) = -pdV \tag{26}$$

leading to :

$$\dot{\rho} = -3(\frac{p}{c^2} + \rho)\frac{\dot{a}}{a} \tag{27}$$

From these two equations, the pressure can be eliminated, and, after having multiply both terms by $\dot{a}$, the differential equation can be easily integrated. This leads to the following equation :

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho}{3} - \frac{Kc^2}{a^2(t)} \tag{28}$$

The last term corresponds to the constant of integration. Its value cannot be specified, depending on the initial conditions. The form of the above equation is independent of the radius a of the sphere

[3]Apparently, this theorem should be named Birkhoff-Jebsen, as it has been published two years earlier by an Norwegian physicist, J.T. Jebsen[13].

and the solution $a(t)$ should be proportional to the quantity $R(t)$. $R(t)$ should also be solution of an equation of the same form, the constant K, which depends on the radius a_0, being related to the constant k which is involved in the Robertson-Walker metric element, something which can be established only within GR :

$$\left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi G\rho}{3} - \frac{kc^2}{R^2(t)} \tag{29}$$

This relation, the Friedamnn-Lemaître equation, expresses the link within the framework of GR between the geometry and the material content of the Universe. In order to specify completely the function $R(t)$, one needs an equation of state for the content of the Universe. The three cases often seen in cosmology are the dust case ($p = 0$), the radiation dominated regime ($p = \frac{1}{3}\rho c^2$) and the cosmological constant equivalent to a vacuum contribution: $p_v = \frac{1}{3}\rho_v c^2$. The cosmological constant Λ being $8\pi G\rho_v$. So the above Friedamnn-Lemaître (FL) equation can be written:

$$H^2 = \left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi G\rho}{3} - \frac{kc^2}{R^2(t)} + \frac{\Lambda}{3} \tag{30}$$

The cosmological parameters are introduced from this expression. The cosmological matter density parameter:

$$\Omega_m = \frac{8\pi G\rho_m}{3H^2} \tag{31}$$

the reduced cosmological constant

$$\Omega_\Lambda = \frac{8\pi G\Lambda}{3H^2} \tag{32}$$

and the cosmological curvature parameter:

$$\Omega_c = -\frac{kc^2}{R^2(t)H^2} \tag{33}$$

I find personaly unfurtunate this convention, as $\Omega_c < 0$ for a positively curved space. The FL equation now reads:

$$\Omega_m + \Omega_\Lambda + \Omega_c = 1 \tag{34}$$

Quantities estimated at present epoch are labelled by 0. For instance the Hubble constant is H_0. This could leads to some ambiguity (Ω_0 refers generally to the matter density).

3 Cosmological tests

The determination of cosmologcial parameters has been one of the most important objectives of cosmologists after the discovery of the expansion of the universe by Hubble. It is importan to provide a precise framework but also in the perspective to test the model. This problematic has become specially important after the theory of inflation which predicted the universe to be flat, something which was firt interpret as a prediction for $\Omega_m = 1$. Although Peebles noticed that the actual prediction of inflation was $\Omega_m + \Omega_\Lambda = 1$, little attention was paid to the cosmologcial constant until the detection of small scale fluctuation in the microwave sky. Indeed these measurements were in agreement with a flat universe and inconsistent with open cosmologcial models with $\Omega_m \sim 0.3$. However the evidence for acceleration as obtained from the Hubble diagram of distant supernovae has been the observational evidence that has lead to a rapid change of paradigm. Since that time the improvment in the accuracy on the estimations of cosmologcial parmeters has been dramatic.

3.1 Geometrical tests

Most cosmological tests in Cosmology, at least those on which present day constraints are based on are geometrical in nature. Therefore they are essentially based on the relation between the coordinate r and the redshift, the general mattig's relation:

$$r = S_k\left(\int_0^z \frac{cdt}{R(t)}\right) = S_k\left(\frac{c}{R_0}\int_0^z \frac{dz}{H(z)}\right) \tag{35}$$

while the dependence of $R(t)$ or $H(z)$ on the cosmological parmeters is given by the FL equation.

The Hubble diagram of distant Supernovae The Hubble diagram was the first geometrical test of relativistic cosmology. Extension of the Hubble diagram to high redshift has been made possible thanks to the use of type Ia Supernovae (SNIa). SNIa at their maximum luminosity ($M \sim -19.5$) reach a luminosity comparable to that of an entire galaxy. This means tha these bright objects can be detected extremely far away. Their are therefore observed as there were in an epoch substantially younger than the present universe. Furthermore there is a relation between the decline rate and the intrinsic luminosity making them suitable for distance measurements at cosmological scale. Because SNIa are rare, large sky area have to be surveyed on a regular basis to collect samples of SNIa. At the end of last century, two groups have independently investigated the distant SNIa Hubble diagram and concluded that supernovae at redshift ~ 0.5 were dimmer by ~ 0.2 mag compared to what was expected in a unaccelerated universe. This was interpreted as an evidence for an accelerated expansion. Indeed as supernova are observed in the universe when younger they allowed to measure the history of the expansion. The consequence is very dramatic: gravity is repulsive on the scale of the universe is accordingly to this observation!

What if Supernovae evolved Given the importance of the consequence not only for cosmology but also for fundamental physics, the above observatin should be scrutinized. The use of geometrical tests is based most of time on the assumption of no-evolution of the parent population. This is also the case for type Ia supernovae. Although strong efforts have been done by observers to track for any sign of evolution by close inspection of the spectra [2], the absence of evidence cannot be considered as an evidence of absence. One possible way to deal with this problem is to assume some evolution and see whether the data still provide evidence for the claim. For instance, an evolution term like:

$$\Delta m_e \propto z \tag{36}$$

can not mimic the observed Hubble diagram without a cosmological constant. However an other form of the evolution term has been suggested, being proportional to the look back time :

$$\Delta m_e \propto \Delta t \tag{37}$$

[29]. It happens that such term leads to large degeneracy between cosmology and possible evolution [11] that present day data do not allow to disentangle .
Undoubtfully, despite its possible limitation, the determination of the Hubble diagram from SNIa has led to a major and rapid change of paradigm in modern cosmology. However, this change has been possible because the previous situation was problematic. Although some observational indications were favoring a low density universe, the first detections of fluctuations on degree scales were in conflict with open low density universe [17].

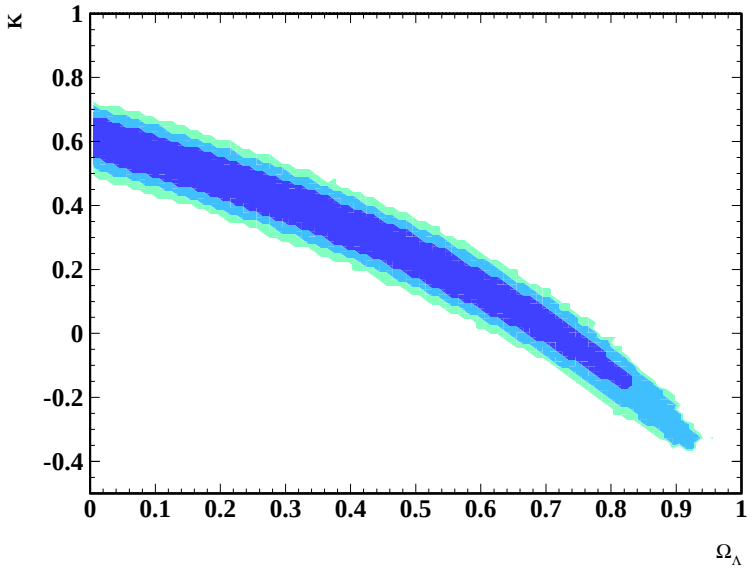


Figure 1: Fitting the SNIa Hubble diagram with two free parameters, one being the cosmological constant in a flat cosmological model and the second being a parameter describing a possible time evolution of the luminosity of distant supernovae ($\Delta m(z) = K(t_0 - t(z))/(t_0 - t(1))$) leads to the following constraints [11]. Contours are 1, 2 and 3 sigma regions. This is a strong degeneracy between the two parameters which prevents an unambiguous evidence for a cosmological constant from the sole Hubble diagram of SNIa.

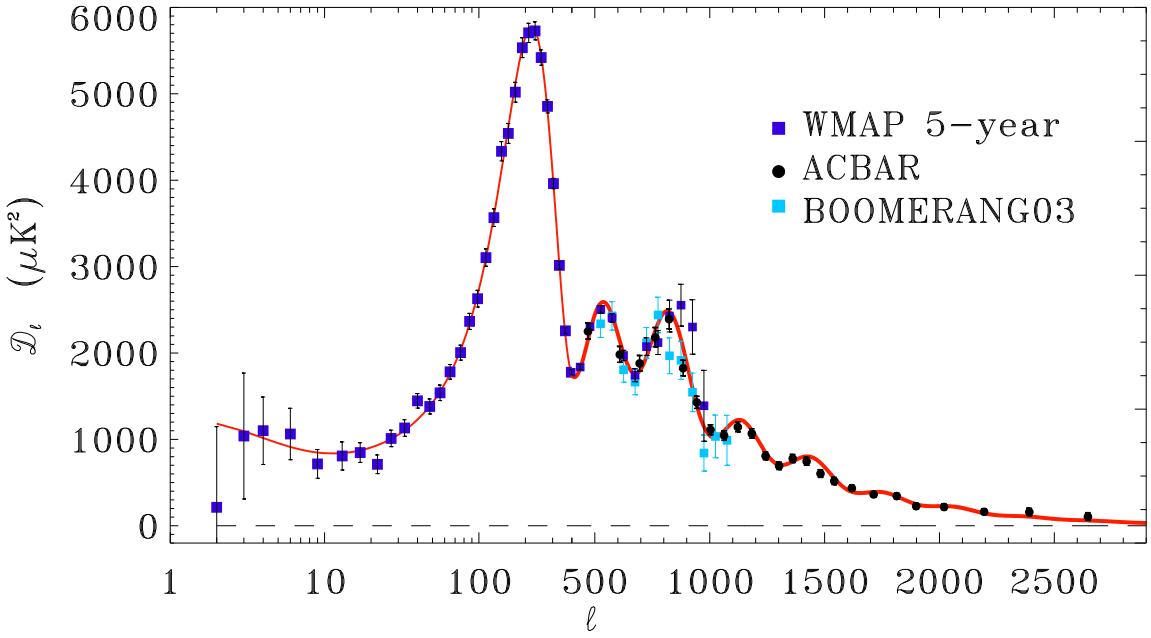


Figure 2: The amplitude of angular fluctuations of the CMB is expressed through their angular power spectrum. Data are WMAP, Boomerang, ACBAR [24]. A simple minimal six parameters model including a cosmological constant provides an excellent fit to the data. This is one of the most important successes of modern cosmology.

Fluctuations Since the discovery of the CMB fluctuations by COBE [26] the idea that early universe physics has left imprints revealed by these fluctuations has gained an enormous attention. In this respect, DMR results have played a fundamental role in modern cosmology comparable to the discovery of the expansion of the universe or the discovery of the microwave background by Penzias and Wilson, and indeed this has motivated the delivering of the Nobel prize to G. Smoot and J. Mather for this discovery. One of the fundamental reasons for this is that fluctuations on scales larger than one degree in the microwave background radiation correspond to scales greater than the horizon at last scatering epoch and cannot therefore been altered by any physical process and should therefore reflect primodrial fluctuations [28]. This also means that the very existence of these fluctuations could be explained only from yet undiscovered physics, probably relevant to the very early universe [18], for which the expansion law is strongly modified compared to the standard picture. The DMR results were providing some constraints on cosmological models [30] but it has been realized that the measure of fluctuations on smaller scales will provide much stringent information. Early detection of fluctuations on degree scales allowed to set interesting constraints and provide the first evidence for a flat geometry of space [16, 17]. If estimations of low matter density were to be regarded as robust, this was inevitabily leading to a non-zero cosmologcial constant. Even before the availability of the WMAP data, considerable progresses have been achieved on the measurement of fluctuations on all angular scales. Archeops [3] and Boomerang [5], as well as many other small scale measurements, already provided data allowing tight constraints on cosmological parameters [4]. It should also be noticed that fast codes to compute the fluctuatiosn spectrum has been made available to the scientific community. The first one was CMBFAST [25] followed by an avatar, CAMB[15]. The authors deserve the warm aknowledgments of the community as these tools have been really critical in the full scientific exploitation of the various CMB experiments.

Although the observed fluctuations were consistent with a Λ dominated universe, a cosmological constant was not explicitly requested by the CMB data alone. Indeed even the WMAP data were consistent with a vanishing cosmological constant, provided the Hubble constant was left as an entirely free parameter. A positive detection of a cosmological constant could be obtained only by using some additional data in conjunction with CMB, like the measurement of the Hubble constant. A further restriction came from the fact that the constraints on cosmological parameters were obtained within the standard CDM picture, and that many ingredients were specified without being necessarily confirmed by observations : for instance initial fluctuations are supposed to be adiabatic and to follow some power law. Therefore the "concordance" [20] cosmology was an appropriate terminology: the model was consistent with most existing data, but the introduction of a cosmological constant was not requested by any single data, and it was far from being clear whether relaxing some of the input hypotheses would not allow for solutions without the introduction of a cosmological constant.

What do actually fluctuations tell ? The first point to notice is that for a random function on the sphere, even with gaussian statistics, each a_l is a random quantity. Therefore fitting the $\mathcal{C}_l$ with an acceptable goodness of fit figure means that several thousands of random numbers could be fitted with a 6-parameter theory. A remarkable level of achievement! In addition fitting the $\mathcal{C}_l$ curve provide very tight constraints on the six parameters, due to the quality of the measurements. These constraints are generally formulated in term of cosmological parameters and it is often quoted that they provide a direct evidence for an accelerating universe independent of the Hubble diagram of supernovae. It should be realized however, that these constraints are established within a specific

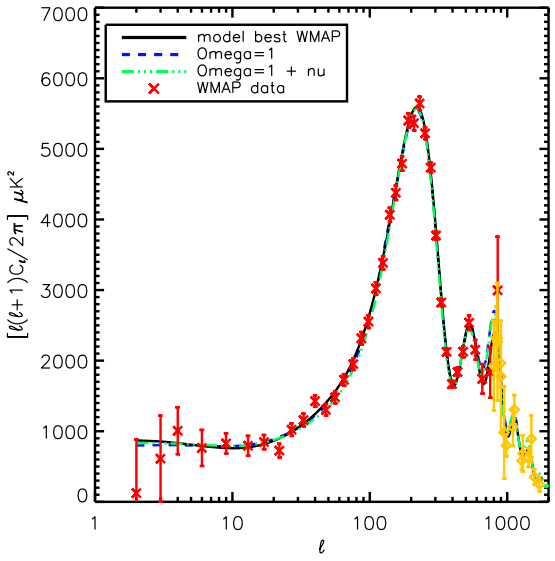


Figure 3: The TT spectrum of the first year WMAP data compared to three different models: one is the concordance, the two others are Einstein de Sitter models, one of which comprises neutrino contribution of $\sim 10\%$ corresponding to three degenerate families with $m_\nu \sim 0.7$eV. From Blanchard *et al.* [7].

model that is the adiabatic cold Dark matter picture with power law initial conditions. Therefore these constraints are model dependent. Modifying the starting hypothesis may change these constraints (and the model may then be rejected, like the standard topological defects scenario has been). An early illustration of this has been obtained soon after the publication of the WMAP data. Relaxing the powerlaw hypothesis, i.e. assuming a non power law power spectrum, it is possible to produce $\mathcal{C}_l$ curves within an Einstein de Sitter cosmological model which provided a fit as good as the concordance model. This is illustrated in figure 3 on which 3 models are compared to the WMAP data, two being Einstein de Sitter models. Such models not only reproduce the TT (temperature-temperature) spectrum, but are also extremely close in terms of ET (polarization-temperature) and EE (polarization-polarization) spectra. An un-clustered component of matter like a neutrino contribution or a quintessence field with $w \sim 0$ is necessary to obtain an acceptable amplitude of matter fluctuations on clusters scales [7]. Such models require a low Hubble constant ~ 46 km/s/Mpc at odd with canonical HST key program value (~ 72 km/s/Mpc) but is actually only $\sim 3\sigma$ away from this value, this can certainly not be considered as a fatal problem for an Einstein-de Sitter universe. The introduction of a non-power law power spectrum might appear as unnatural. However, such a feature can be produced by some models of inflation in order to match the $\mathcal{C}_l$ curve [12]. Therefore the amplitude and shape of the CMB fluctuations as measured by WMAP is certainly a success for the Λ CDM model but cannot be regarded as a direct indication of the presence of dark energy.

Large scale structure Within a specific model like Cold Dark Matter, not only it is possible to derive the $\mathcal{C}_l$ curves, that is the angular power of the fluctuations of the cosmic microwave background, but it is also possible to obtain the power spectrum of the fluctuation in the matter density, or equivalently the correlation function. The galaxy distribution should reflect essentially this matter power spectrum (galaxies may be a "biased" representation and this bias is subject to some modeling, but this represent small perturbations). The measure of the power spectrum can therefore be used to disentangle models which produce C_l curves that could not be distinguished. Recently, a critical advance resulted from the availability of very large galaxy surveys, the 2Df redshift survey and the SDSS survey, allowing to measure the amplitude of galaxy fluctuations on scales as large as $100h^{-1}$ Mpc [22, 27, 10, 23]. This has provided a remarkable success to the ΛCDM picture because the shape of the correlation function could be predicted for models that already match the CMB fluctuations measured by WMAP: not only ΛCDM model reproduces the shape of the correlation function, but the specific presence of a bump in the correlation function at scale of the order of $100h^{-1}$Mpc due to the detailled dynamics of fluctuations when the baryons are taken into account, the so called accoustic peak, corresponding to the "peak" in the $\mathcal{C}_l$ of the CMB.

Once an Einstein de Sitter model is built in order to reproduce the CMB $\mathcal{C}_l$, the amplitude of the matter fluctuations on large scales is set up and the measurement of the matter fluctuations on large scales in the present day universe is a critical way to distinguish models which are otherwise degenerated in their $\mathcal{C}_l$. The comparison of the power spectrum from the SDSS LRG with the predicted spectra for Einstein de Sitter models is clearly in favor of the concordance model, see Fig. 4. One should add some caution here: it might be possible that the biasing mechanism leads to a power spectrum at small k (large scales) which is not proportional to the actual matter power spectrum [9], in which case the above comparison might not be a fatal failure of the Einstein de Sitter models. However, biasing mechanisms systematically lead to a correlation function on large scales which is still proportional to the matter correlation function on large scales. Comparison of the correlation function on large scales is therefore less ambiguous and its measurement should

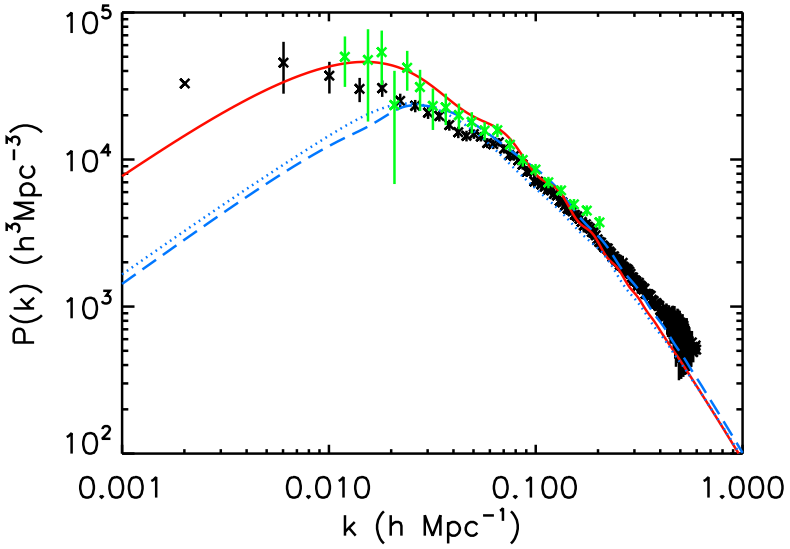


Figure 4: Data from the SDSS have allowed to measure the amplitude of galaxy fluctuations on large scales. In this respect, Luminous Red Galaxies (LRG) provided measurement of the power spectrum on the largest scales. Green crosses correspond to Tegmark et al. [27] and black crosses correspond to the measurements of the power spectrum of LRG from the SDSS Data Release 5 by Percival et al. [23]. The red continuous curve is the predicted spectrum for a typical concordance model, while the dotted and dashed lines correspond to the power spectrum for Einstein de Sitter models consistent with the WMAP fluctuation angular power spectrum $\mathcal{C}_l$ [7, 12].

be unambiguously discriminant. Hunt and Sarkar [12] have provided a comprehensive MCMC investigation of the Einstein de Sitter parameter space, finding models which acceptably fit the correlation function on scales below 70 h^{-1}Mpc, but were nevertheless systematically negative on scales of the BAO peak. This is a strong evidence that there is no way in an Einstein de Sitter universe to fit simultaneously the $\mathcal{C}_l$ and the observed distribution of galaxies on large scales. This should be regarded as a remarkable success of the concordance cosmological model: although there were little doubts that this model could fit accurately most of the major existing observational facts in cosmology, the ability to produce predictions that are verified a posteriori is the signature of a satisfying scientific theory.

Parameter	**Vanilla**	**Vanilla** $+ \Omega_k$	**Vanilla + w**	**Vanilla** $+ \Omega_k + w$
$\Omega_b h^2$	0.0227 ± 0.0005	0.0227 ± 0.0006	0.0228 ± 0.0006	0.0227 ± 0.0005
$\Omega_c h^2$	0.112 ± 0.003	0.109 ± 0.005	0.109 ± 0.005	0.109 ± 0.005
θ	1.042 ± 0.003	1.042 ± 0.003	1.042 ± 0.003	1.042 ± 0.003
τ	0.085 ± 0.017	0.088 ± 0.017	0.087 ± 0.017	0.088 ± 0.017
n_s	0.963 ± 0.012	0.964 ± 0.013	0.967 ± 0.014	0.964 ± 0.014
Ω_k	0	-0.005 ± 0.007	0	-0.005 ± 0.0121
w	-1	-1	-0.965 ± 0.056	-1.003 ± 0.102
Ω_λ	0.738 ± 0.015	0.735 ± 0.016	0.739 ± 0.014	0.733 ± 0.020
Age	13.7 ± 0.1	13.9 ± 0.4	13.7 ± 0.1	13.9 ± 0.6
Ω_M	0.262 ± 0.015	0.270 ± 0.019	0.261 ± 0.020	0.272 ± 0.029
σ_8	0.806 ± 0.023	0.791 ± 0.030	0.816 ± 0.014	0.788 ± 0.042
z_{re}	10.9 ± 1.4	11.0 ± 1.5	11.0 ± 1.5	11.0 ± 1.4
h	0.716 ± 0.014	0.699 ± 0.028	0.713 ± 0.015	0.698 ± 0.037

Table 1: Summary of the mean values and 68% confidence intervals for the cosmological parameters of the ΛCDM model constrained from CMB, SNIa and BAO for different models (θ is the ratio of sound horizon to angular diameter distance). These constraints are quite tight, most of them are below 5%, and are stable when additional degrees of freedom are added to the model (w, Ω_k), adapted from [11].

3.2 Tests based on the growing rate

As we have seen the Cold Dark Matter should be regarded as a successful theory that has lead to predictions which were verified a posteriori and which is able to reproduce most of the data relevant to cosmology. The precision on cosmological parameters for the ΛCDM picture is of the order of 5% at most, with accuracy close to 1% in some case. Of course this doesn't mean it is the "right" theory; science does not provide "right" theories but only theories that reproduce all existing data and which are able to lead to predictions that can lead to its *invalidation*. This is the principle that to be scientific a statement has to be falsifiable accordingly to Kark Popper.

It is therefore the only possible path to continue to increase the accuracy of existing measurements and to develop new ways to test the theory. There is a way to test cosmological models which is fundamentally different from geometrical tests: it is based on the growing rate of fluctuations under their own gravity. In principle the abundance of clusters and weak lensing measurements are both sensitive to this growing rate. I do not think that they have by now reach a level of precision

that makes them useful, but it is certainly a way on which efforts will concentrated in the future, in particular thanks to space mission like EUCLID or WFIRST. Measurements of this growing rate would allow to test whether predictions of general relativity at the scale of the universe are verified or if we have to turn to alteratives.

4 Conclusions

The Copernician model of the world was the first revolution of a series in the construction of modern cosmology, and the discovery of the accelerated expansion being the latest in date. Theoretical considerations have always been a source of remarkable observational investigations and Cosmology has always benefited from the confrontation of models with observations. Since the thirties, the big bang picture, the modern version of Lemaître's primeval atom has been remarkably successful, based on simple assumptions and physical laws that have been validated by accurate experimental results. Although alternative theories have been developed, these alternative were based on hypothetical unknown physics advocated to interpret cosmological observations. None of these alternative theories has produced significant predictions differing from the standard view that would have been comforted a posteriori. Rather new observations in agreement with predictions of the big bang picture necessitated deep revision of the unorthodox views, at the cost of rather ad hoc assumptions added to fit the new observations. The situation has evolved when the standard picture has necessitated the introduction of new ingredients, first dark matter and more recently dark energy. The very nature of these new ingredients, which are supposed to dominate the mean density of the universe has not been established by direct laboratory experiments, nor by astronomical observations, and this situation may some time lead to the question whether cosmologists have not introduced new aethers. We had the opportunity to see that the situation is not so bad. The introduction of -cold- non-baryonic dark matter has led to specific predictions, the amplitude and shape of the fluctuations of the cosmological background on various angular scales, which were verified with high accuracy. The presence of dark energy has lead to a specific prediction, the shape of the matter power spectrum on large scales, which has been verified a posteriori. Although the inclusion of a cosmological constant was concomitant to general relativity, the actual origin of dark energy remains totally unknown and the presence of dark energy in the present day universe represents probably the most fundamental and unexpected new element in modern physics.

References

[1] Abramowicz, M. A., Bajtlik, S., Lasota, J.-P., & Moudens, A., *Acta Astronomica*, **57**, 139 (2007).

[2] Balland, C., et al., *A&A*, **507**, 85 (2009)

[3] Benoît, A., et al., *A&A*, **399**, L19 (2003).

[4] Benoît, A., et al., *A&A*, **399**, L25 (2003).

[5] de Bernardis, P., et al., *Nature*, **404**, 955 (2000).

[6] Birkhoff, G. D., Relativity and Modern Physics, Harvard University Press (1923).

[7] Blanchard, A., Douspis, M., Rowan-Robinson, M., & Sarkar, S., *A&A*, **412**, 35 (2003).

[8] Cook, R. J., & Burns, M. S., *American Journal of Physics*, **77**, 59 (2009).

[9] Durrer, R., Gabrielli, A., Joyce, M., & Sylos Labini, F., *ApJL*, **585**, L1 (2003).

[10] Eisenstein, D. J., et al., *ApJ*, **633**, 560 (2005).

[11] Ferramacho, L. D., Blanchard, A., & Zolnierowski, Y., *A&A*, **499**, 21 (2009).

[12] Hunt, P., & Sarkar, S., *Phys. Rev. D*, **76**, 123504 (2007).

[13] Johansen, N. V., & Ravndal, F., *Gen. Rel. and Grav.*, **38**, 537 (2006).

[14] Lachieze-Rey, M., & Luminet, J., *Phys. Rep.*, **254**, 135 (1995).

[15] Lewis, A., Challinor, A., & Lasenby, A., http://camb.info/, *ApJ*, **538**, 473 (2000).

[16] Lineweaver, C. H., Barbosa, D., Blanchard, A., & Bartlett, J. G., *A&A*, **322**, 365 (1997).

[17] Lineweaver, C. H., & Barbosa, D., *ApJ*, **496**, 624 (1998).

[18] Guth, A. H., *Phys. Rev. D*, **23**, 347 (1981).

[19] Luminet, J.-P., Weeks, J. R., Riazuelo, A., Lehoucq, R., & Uzan, J.-P., *Nature*, **425**, 593 (2003).

[20] Ostriker, J. P., & Steinhardt, P. J., *Nature*, 377, 600 (1995).

[21] Peacock, J. A., arXiv:0809.4573 (2008).

[22] Percival, W. J., et al., *MNRAS*, **327**, 1297 (2001).

[23] Percival, W. J., et al., *ApJ*, **657**, 645 (2007).

[24] Reichardt, C. L., et al., *ApJ*, **694**, 1200 (2009).

[25] Seljak, U., & Zaldarriaga, M. *ApJ*, **469**, 437 (1996).

[26] Smoot, G. F., et al., *ApJL*, 396, L1 (1992).

[27] Tegmark, M., et al., *Phys. Rev. D*, **74**, 123507 (2006).

[28] Weinberg, S., Gravitation and Cosmology, ISBN 0-471-92567-5. Wiley-VCH (1972).

[29] Wright, E. L., arXiv:astro-ph/0201196 (2002).

[30] Wright, E. L., et al., *ApJL*, **396**, L13 (1992).

Type Ia Supernovae and the discovery of the Cosmic Acceleration

Alejandro Clocchiatti
Departamento de Astronomía y Astrofísica
Facultad de Física
Pontificia Universidad Católica de Chile
&
High Z Supernova Search Team

October 11, 2011

Abstract

I present a review of the research and analysis paths that converged to make Type Ia SNe the most mature cosmological distance estimator of the present time. The narrative starts with the first works in the early decades of the 20th century and finishes with the more recent results, covering the surprising discovery of Cosmic Acceleration in 1998.

The review was written thinking of physicists with a strong interest in Cosmology, who might have pondered why was that, after decades of not being able to agree upon the rate of cosmic expansion, astronomers were so quick to concur on cosmic acceleration.

1 Introduction

Supernovae (SNe) have been with us from the beginning, since before we were "us". Occasionally superimposed on the background of a familiar, regular and repetitive sky, they were among the variable phenomena that puzzled, marveled and scared our human ancestors. The modern view associates them with the final stages of stellar evolution in scenarios where the shifting balance between pressure sources and gravitation reaches unstable regimes. When the instabilities lead to an explosion powerful enough to disrupt the star we have SNe.

The current paradigms purport SNe as critical ingredients in the micro-physics of Cosmology. They create the chemical history of the Universe, stir the interstellar medium of galaxies enriching primordial matter with novel chemical elements, and, if exploding in dense environments, generate shock waves that promote the birth of newer generations of stars. In addition, from the practical point of view of today′s astronomers, they provide outstanding estimators for cosmologically relevant distances.

This paper is a review of the research paths that converged to make Type Ia SNe, at the time, the most mature distance estimator for use in Cosmology, the surprising discovery of Cosmic Acceleration in 1998, and touches the current work trying to use SNe to obtain even finer cosmological inferences. Opposite to what the organizers of the meeting expected, the paper was written not

before, but *after* the presentation at the workshop. The positive side of having done it this way, is that I could tune it to answer the questions I received after the talk. It was the first time I presented the subject to an audience made up, primarily, of physicists, and the questions were different to what I was typically used to answer. Hence, I prepared this work thinking mainly of colleagues who are physicist that work in Cosmology. I will touch some topics considered "basic" in SNe studies, which are available in other reviews or papers, but are scattered around mainly in astronomical journals. My hope is to present basic concepts, give an overview of the advances as they proceeded in time, and organize some relevant bibliographic sources to facilitate further study from scholars who are not specialist in the field.

I must warn the reader that the text, as a review of SNe, is not complete. It cannot be. It is biased towards the developments that lead Type Ia SNe to become the outstanding distance estimators that they are today. The rest of the SNe appear and are mentioned just as the background out of which Type Ia SNe appeared and were recognized.

I shall also warn the reader that I have made an attempt to merge in this manuscript the different styles of referencing literature used by physicists, like in Phillips [69], and astrophysicists, like in Phillips (1993). Being myself an astrophysicist I have come to appreciate the importance of having the date of the papers inserted in the text, to figure out the pace of developments and/or possible historical relation between different pieces of work. The result, Phillips (1993 [69]) is slightly cumbersome, but I hope that will suite the prejudices and slants of both communities.

The chart of this paper is as follows. Section 2 is a historical narrative. In it I describe the evolution of the area that I loosely call "SN Studies," from the times of the pioneers, in the early decades of the 20th Century until all the major advances in instrumentation, observation, analysis, and theoretical interpretation had taken place, leaving the field poised for a major breakthrough by 1993. In Section 3, I describe the fast developments between 1994 and 1998, when the Cosmic Acceleration was discovered. In Section 4, I present the results of 1998, using as an illustration the complete set of Type Ia SNe published by the High Z SN Search Team, augmented by selected later discoveries. In Section 5, I describe what could be called the "Second Generation" SN surveys started soon after the turn of the century, with the goal of constraining the nature of the now called "Dark Energy".

2 Supernova Studies

2.1 The pioneers

The serious study of SNe started with two very influential papers by astronomers from the Carnegie Institution and California Institute of Technology. Baade and Zwicky (1934a [8] & 1934b [9]), proved clearly that the nova stars known from antiquity were composed by two very different classes. One of them, which they named for the first time super-novae, was much more powerful than the other. They showed that the energy involved in these outbursts was equivalent to a considerable fraction of the rest mass of a star and they proposed that the neutronization of matter could be the source of energy for the process. They issued the suggestion as a general possibility based on energy considerations as they did not present any specific scenario where neutronization could occur. Specifically they did not focus on the, now preferred, gravitational collapse in the core of massive stars. They also suggested that SNe could be one of the sources of Cosmic Rays, a possibility that has been explored and confirmed since.

Having associated supernovae with the final stages of the still largely unknown stellar evolution

it was natural to device a SN search program. If a survey is done, imaging nearby galaxies with hundreds to thousands of millions of stars each, giving enough time, chances are that a few of those stars will be caught in the act of becoming SNe. Once they are discovered, a systematic program for studying the outbursts could be done just by following the evolution in time of the emitted flux and spectrum. Using the recently commissioned wide angle 18-inch Schmidt telescope at Palomar observatory, Zwicky, together with J. Johnson, started the first systematic search for SNe in 1936. The telescope would be used to image hundreds of galaxies in multiple epochs during the observing season and compare the new images with the old ones using blinking microscopes. Any new star could be easily spotted. Zwicky teamed up with another astronomer of the Carnegie Institution to follow up the spectroscopic evolution. After a few years of work, Minkowski (1941 [51]) was able to start the dissection of the SN events classifying them into "types" according with their spectra, a tendency that has only partially helped to clarify the field, but, in any case, has been impossible to change.

Minkowski proposed that SNe come in two different spectroscopic types, I and II. The early time spectra of Type I SNe were not understood. They showed wide absorption/emission bands that shifted in position and strength with the passage of time. But it was not possible to identify the chemical elements that caused them. It was clear, though, that they were a homogeneous class because the unknown spectra would repeat from one event to the next. A few spectra taken hundreds of days after outburst provided broad features that could be reasonably associated with emission lines, out of which only one forbidden transition of O I was identified. The evolution in time was also remarkable similar for different SNe within this type. Type II SNe, on the other hand, displayed the well known lines of the hydrogen Balmer series and the He I line at 5678 angstroms. It was easy to recognize the characteristic shape of P Cygni profiles and, hence, measure expansion velocities. These turned out to be in the order of several thousand kilometers per second. Thus, the hypothesis of an explosion with massive ejection of matter was confirmed for Type II SNe and, by extension, adopted as a reasonably working hypothesis for Type I as well.

Zwicky also contributed to the initial proliferation of SN types by using the characteristics of the photographic light curves (the light curve is the evolution of brightness in a given photometric band with time) as a tool to aggregate SNe into classes. This idea also stuck, although the groups that he proposed have been abandoned (see for example a later recollection in Zwicky, 1965 [101]).

Classification of events aside, the photometric follow up provided a puzzling result. The SNe sustain the emission of energy for hundreds of days. Type I, in particular, after they go through the maximum emission of light during the initial weeks, settle on a linear decay of brightness with time if plotted as magnitude (i.e. logarithm of the emitted energy). This exponential decay phase was followed for hundreds of days in some events.

As the database of observations increased, the spectra of Type I SNe continued to be a source of perplexity. It was not even known how to interpret them in the most basic terms. Were they emission line spectra of atoms in different ionization states whose lines were enormously broadened by large expansion velocities? Were they combined spectra, where a continuum and broadened absorption and emission lines are present, like an extreme case of some peculiar stars or normal novae? The first serious attempt to make sense of Type Ia spectra was by Payne-Gaposchkin and Whipple (1940 [60]) and Whipple and Payne-Gaposchkin (1941 [97]). They attempted to reconstruct the SN spectra by summation of the emission lines of the more common chemical elements in astrophysics, allowing for several stages of ionization. They produced synthetic spectra that were a very rough match to the observations and left open the possibility that many of the troughs in the spectra were real absorption minima that they had not tried to model.

The pioneering work described above created the field of SN studies and set the landscape and main inroads for its subsequent evolution. The early discoveries were puzzling, and left broad and important questions to be answered. The neutronization of matter was an fascinating possibility for the production of the energy required to disrupt the stars, but there were no scenarios identified for such a process to occur. In addition, the fact that there were two SN types suggested that there could be two different explosion mechanisms. The early attempts at theoretical interpretation of SNe spectra explored for the first time what, in the end, would be the key approach to decipher them: the construction of synthetic spectra.

2.2 Getting the picture: The 40s and 50s

The 40s and 50s were a period of slow accumulation of observations and evolution of the conceptual tools to understand, in general, astrophysics. Basic knowledge on atoms, atomic nuclei, nuclear reactions, and improvements in the measurements of the cosmic abundances, helped to establish the cosmological role of stars and stellar nucleosynthesis.

The data of Zwicky's first survey continued to be studied, until they started to reveal its full content. An important event, which will play its role later on, was the discovery and observation of a bright SN in NGC 4214. This SN was observed from Europe and the US and Wellmann (1955 [93]) remarked that it was of Type I, but the spectrum was somehow peculiar (we know now that it was the first Type Ib SN with spectra recorded).

Another relevant piece of observational information that finally came to be recognized was the outstanding light curve of SN 1938C observed by Zwicky, and presented by Baade (1945 [6]). The part of the light curve that comes after maximum, the exponential decay, had a half-life of $\sim$55 days. This slope was essentially the same for all Type I SNe and became a strong argument in favor of a common energy source for all Type Ia SN. For SN 1938C, the brightest Type Ia SN since the invention of the telescope and the better observed event at the time, the exponential decay lasted as long as the observations, more than 10 times the half-life. Borst (1950 [11]) squarely set the focus on this fact and concluded that it was the natural result a radioactive decay. Connecting directly the half-life of the SN light curve with the half-life of the unstable nuclei, he suggested that ^{7}Be was the energy source and proposed a mechanism to form this unstable nucleus following the collapse of the stellar core after hydrogen exhaustion. Baade et al. (1956 [7]) and Burbidge et al (1956 [15]) picked up the idea, emphasizing that in addition to ^{7}Be, both ^{89}Sr, and the recently discovered ^{254}Cf, were also good matches for the half-life time, and they justified their preference for the latter as the energy source for SNe.

Observations of abundance of elements in meteorites, Earth, the Sun, stars and nebulae, also started to converge and it was possible to interpret them in terms of the recently created nuclear shell model. Suess and Urey (1956 [85]) published a paper that set the standard of the time, with the first modern looking plots of the relative abundances of chemical elements.

Most of the ideas and advances of the last decades that were slowly converging to the big picture of astrophysics were organized in one of the more influential papers of the 20th century astrophysics by Burbidge, Burbidge, Fowler and Hoyle (1956 [15], hereinafter named B2FH). In this paper, they presented a coherent picture for the origin of the chemical elements in thermonuclear reactions of increasingly heavier nuclei in the core of evolving stars. They also emphasized the need for additional neutronization of complex nuclei, and developed the concept of slow and rapid neutron capture processes. The slow processes would take place sometime during normal stellar nucleosynthesis and the rapid ones in the explosions of SNe. It was, essentially, the modern view, with a few notable

shortcomings that are easy to pinpoint now, 55 years after. Many of those were related with stellar astrophysics and SNe. In the effort to provide the synthesis authors disregarded the fact that the two spectroscopic SN types could point out to the existence of two different progenitors and/or explosion mechanism. B2FH correctly located the terminal instability of core-collapse SNe in the ^{56}Fe core of massive stars that had followed all the nucleosynthesis sequence, but did not have a clear understanding of the processes that triggers the collapse. They did not have, either, a clear understanding of how the core collapse would result in the ejection of the rest of the star and the explosive nucleosynthesis. Their concept was that the core collapse will start a rapid compression inwards of the whole star, including the mantle. This would sharply raise the temperature of gas of the mantle, rich in nuclear fuels, and trigger a runaway thermonuclear reaction which will disrupt the star. B2FH also insisted on ^{254}Cf as the radioactive product that would power the light curves, at times as if the presence of ^{254}Cf in SNe were an observational fact.

2.3 Getting the picture right: The 60s and 70s

The 60s started well for SN science. Hoyle and Fowler (1960 [43]), recognized the importance of light nuclei (^{12}C, ^{16}O, ^{24}Mg, etc.) as thermonuclear fuel when they were in highly degenerate conditions. They identified degenerate low mass stars near the Chandrasekhar limit as an optimal scenario for explosions sustained by these fuels, and associated these explosions with Type I SNe. They estimated the mass of terminally unstable stars and found it to be constrained to a very narrow range. Out of this fact they explained the relatively scarcity of the explosions, but failed to recognize the, even more remarkable, relevance for the homogeneity of the class. This is the first paper where collapse of evolved cores and thermonuclear explosions are clearly associated with Type II and Type I SNe, respectively. Hoyle and Fowler revisit as well the ^{254}Cf hypothesis and found it not as compelling as in previous papers, because other radioactive heavy nuclei had been discovered with similar half-lives. Also, for the first time clearly, they stressed that the half-life of the powering nucleus need not be directly matched by the half-life of the SN light curve.

Colgate (see for example Colgate et al. 1961 [21]), Arnett (1966 [3]), and Truran (1966 [90]) started a long series of quantitative studies of spherical shocks running in stellar envelopes, the associated nucleosynthesis, and how their energy output could match the observed SN mass ejection and light curves. An important advance came after Colgate and White (1966) realized that the B2FH paradigm of thermonuclear explosion induced by collapse of the core in massive stars does not work, because the collapse actually starts a rarefaction wave at the inner part of the mantle that quenches thermonuclear reactions. What they found, in turn, was that the dynamical implosion of the evolved core was so violent that a vast amount of energy, many times greater than the thermonuclear one, was available just from the deepening of the gravitational potential well. This energy was transferred to the mantle via the emission and deposition of neutrinos. This was the birth of the "prompt shock" model for core-collapse SN explosions.

Truran, Arnett and Cameron (1967 [91]) studied the process of Si burning. Afterward, Truran himself suggested Colgate to consider the consequences of Si burning in his model light curve calculations. Colgate and McKee (1969 [23]) presented the first numerical results where theoretical light curves are a reasonable match to observations. The importance of the α-particle isotope ^{56}Ni, which decays into ^{56}Co with a half-life of 6.1 d, and then into ^{56}Fe, with a half-life of 77.12 d, had been found. This was conceptually very important to make astronomers realize that essentially the full display of Type I SNe was due to radioactive decay, and not just the exponential tail.

Finzi and Wolf (1967 [28]), picked up the thermonuclear runaway in degenerate matter proposed

by Hoyle and Fowler (1960 [43]). They focused on massive white dwarfs as likely candidates for a catastrophic change that would trigger the explosion, thought of those SNe that have been observed in elliptical galaxies, where star formation ended more than 10^{10} years ago, and asked the right question: How can a star that reached the white dwarf stage so long ago suddenly become a Type I SN? They went on trying to answer it assuming that very slow electron captures dynamically change the dwarf star and trigger collapse, heating, and thermonuclear runaway. For the process to work, they had to build white dwarfs of fairly exotic composition. Hansen and Wheeler (1969 [39]) took the idea a step further and numerically computed the explosion of a white dwarf of ^{12}C. They showed that the collapse and nuclear detonation was sufficiently catastrophic so as to explode the whole star providing energies and ejection velocities that matched the observations. Wheeler and Hansen (1971 [94]) expanded the calculation to ^{12}C and/or ^{16}O and computed the nucleosynthesis products. They found that the whole nuclear fuels were burnt all the way to nuclear statistics equilibrium (i.e. all iron group elements). They also proposed the model of a white dwarf close to the Chandrasekhar mass accreting matter from a close binary companion as a possible scenario for Type I SNe. Truran and Cameron (1971 [92]) seem to have reached the idea independently, and they went on to postulate that ^{4}He ignition in the accreted matter will trigger the ^{12}C runaway inwards, and then the explosion (this is still considered a possible explosion mechanism for some Type Ia SNe). Arnett (1969 [4]) also explored numerically the detonation of a ^{12}C white dwarf, or white dwarf-like core in an intermediate mass star. A well as Hansen and Wheeler in the previous papers, he found that it created too much iron group elements, a fact that was probably inconsistent with the chemical history of the Galaxy. He concluded that not many SNe could be of this type. Arnett realized, however, that the nucleosynthesis products were very sensitive to the critical density at which ^{12}C ignites, and stressed that a lower value of this density would help to produce elements between ^{12}C and ^{56}Ni, instead of just iron peak ones. Nomoto, Sugimoto and Neo (1976) asked the critical question: Who knows that the detonation is actually initiated [in a C-O white dwarf]? They proposed that the thermonuclear combustion could take the form of a subsonic deflagration, develop convective instabilities, evolve violently outwards and anyway disrupt the star, without ever entering the detonation regime. They found that, if the deflagration is slow, the combustion proceeds in two phases, the second of which starts when the star has expanded. This second phase takes place at a lower density and avoids overproduction of the iron peak elements.

Regarding the understanding of SN spectra, McLaughlin (1963 [49]), an expert on Novae, revisited the spectroscopic plates of SN 1954A that had been taken at Lick Observatory and mostly neglected afterward. He was interested by the remark from Wellmann (1955 [93]) that the SN was "peculiar", and decided to make his own line identifications. After carefully looking at all the recorded spectra he concluded that they were combined emission and absorption spectra. After coming up with the concept of absorption-like and emission-like features, he thought that some spectra looked similar to those of B stars, with little or no hydrogen. He recognized that a couple of absorption-like minima in the blue region, if identified with He I lines, gave consistent velocities, and went on to identify many other features all shifted by velocities of $\sim$5000 km s^{-1}, and broadened by velocities of order 10^3 km s^{-1}. Minkowski (1963 [52]), focusing mostly in the identification of emission-like features, criticized the approach. Pskovskii (1969 [73]) recognized the value of both the realization of the dual (absorption and emission) character of the spectrum by McLaughlin, and the stress of Minkowski on how uncertain was the identification of emission-like features. He realized that the important issue was that, finally, *absorption lines had actually been identified upon a continuum*, and applied the idea to the still unknown spectra of the normal Type I SNe. He struck gold. His paper reveals for the first time that Type I SNe display low excitation lines of low ionized

species, with no trace of light elements. He first correctly identified the, now, characteristic lines of Si II, and then went on to find Fe II, Mg II, Ca II, and S II. This pattern of lines has hold since then. His, more doubtful, identification of He I, has been superseded. Mustel (1971a [53], 1971b [54], 1972 [55]), provided some complementary identifications, and some differing interpretation. Branch and Patchett (1973 [14]) settled the issue in a very elegant way, developing the first modern synthetic spectra and comparing them with the better observed ones.

It was in this period, as well, that SNe were first identified as valuable light sources for cosmological estimation of luminosity distances. Kowal (1968 [48]) collected the first sample of 22 well observed Type I SNe with light curves in photographic magnitudes, found that they had a scatter of $\sim$0.6 mag at maximum, and proposed a program to find SNe in clusters of galaxies with the goal of reducing the uncertainty up to 0.1 or 0.2 magnitudes by averaging many of them. He pointed out an important fact. Type Ia SNe were powerful sources of light but of stellar origin. As such, they were very probable not subject to the strong evolutionary effects that systematically biases the luminosity distances computed using objects that are aggregates of stars. Both galaxies and clusters of galaxies are evolving very fast with time in the region of the Universe available for observation. But stars have been fairly similar to themselves for thousands of millions of years. He was visionary enough to suggest that when light curves for very distant Type I SNe became available, they could be used to know not only H_0, the Hubble constant, but also "the second-order term in the redshift-magnitude relation" (i.e. the deceleration parameter).

By the end of the 70s, the scenarios to produce Type I SN progenitors and explosions were fairly sophisticated. The match between theory of spectra and observation started to make sense and provided a tension which would not be solved for a few years: Theoretical explosions produced mainly iron group elements, but the observed spectra displayed ^{56}Fe but also appreciable amounts of intermediate mass elements, which were not produced by the theoretical models. On the other hand, Type I SNe had been identified as remarkable candidates for measuring cosmologically relevant distances, providing a strong incentive to programs for their search and study. By the end of the decade both Colgate (1979 [20]) and Tammann (1979 [88]) proposed cosmology with Type I SN as one of the major science drivers for the future Space Telescope.

2.4 Getting the picture in focus: 1980-1993

The early eighties witnessed very fast developments in SN science. Some were triggered by observations. In 1981 a SN was discovered in NGC 4536, it was a Type I event that was followed from McDonald Observatory with good wavelength and time sampling up to more than a hundred days after maximum light. Large telescopes and modern electronic detectors were used, which resulted in the best SN spectra ever obtained. The data was analyzed by Branch et al. (1983 [13]) who produced a paper that set the standard for theorists to compare with. Both spectroscopic observations and synthetic spectra confirmed the picture that the ^{56}Ni $\rightarrow ^{56}$Co$\rightarrow ^{56}$Fe was required, and that, in addition to the radioactive Ni, a sizable fraction of Si, Ca, Na, S, Mg and O were also ejected.

The improvements in detector technology continued to play a role. Elias and collaborators started a program to follow up SNe in the infrared from Cerro Tololo Inter-American Observatory (hereinafter CTIO; Elias et al. 1981 [26]). By 1985 they had found that the infrared eye saw two different types of light curves among SNe of the type I, where the optical view had seen just one. Elias et al (1985 [27]) coined the names "Type Ia" for the more frequent, bluer, and typically brighter SNe and "Type Ib" for the other. Spectroscopy at McDonald Observatory indicated that the two classes had different spectra (Wheeler and Levreault, 1985 [96]). It took few years to identify strong

He I lines in the spectrum of Type Ib SNe (Harkness et al. 1987 [41]), and a few more to realize that there were some peculiar Ia spectra with neither Si II nor He I lines, making it natural to create the Type Ic bin for them. The advances in the field by the end of the decade are reviewed by Wheeler and Harkness (1990 [95]).

The eighties also saw the start of the first SN surveys based on automatic telescopes with digital detectors. Some of them targeted local SNe and some of them were specifically designed to crop distant SN for using as cosmological proves. Some, in addition, planned an automatic pipeline of data reduction and image comparison to detect SNe candidates with minimal human intervention. One of these first initiatives was the Berkeley Automated SN Search (BASS, Kare et al. 1981 [45]). The BASS took many years to fly. Perlmutter (1989 [62]) reports the status after (almost) the first year of real time automatic operation. At about the same time the BASS also started to plan and design a search for more distant SN to use in cosmology (Couch et al. 1991 [25]).

The first group to take the leap, and really go after distant SNe to set constraints on cosmology was a Danish group (Hansen et al. 1987 [40]). The search was based on the 1.5m Danish telescope at ESO La Silla, in Chile. It was already a modern search with CCD cyclic imaging, and real time data analysis including image matching and subtraction. Due, mainly, to the small size of the detector, the impossibility of using the R photometric passband band due to strong fringing (this caused by standard interference in the thin silicon layers of the detectors, and was usual in the older CCDs but greatly improved along the nineties), they discovered only one useful SN in two years and then dropped the program (Noorgard-Nielsen et al. 1989 [59]).

Looking back at their effort is clear that they were way ahead of the proper time. But the finding of SN 1988U at z=0.31 proved that the strategy of multi epoch imaging and digital image comparison did lead to discover distant SNe. By the early nineties the Berkeley SN team made an agreement to put a large CCD camera on the 2.5m Isaac Newton Telescope in the Canary Islands in exchange for observing time to search for distant SNe. It was this combination that allowed them to discover SN 1992bi at z=0.46, the more distant SN up to that time (Perlmutter et al. 1995 [64]). After that success the group started to compete for time at the run of the mill 4.0m class telescopes of the world, and finding distant SN became more usual.

Another critical development of the time was the spread of large CCD detectors around telescopes at different observatories. This made it possible to follow up SN even in non photometric conditions. Since it was typical that the parent galaxy and the SN would appear in a field together with some Galactic foreground stars, relative photometry could be done with respect to a local standard sequence. By the early nineties, good quality, well sampled, multicolor light curves of SNe in nearby galaxies started to accumulate. A compilation by Sandage and Tamman (1993 [83]) found that the scattering of the B maximum brightness of SN Ia, treated as if they were standard candles, around the linear relation implied by the Hubble flow was 0.51 mag. This was a modest improvement from the 0.6 mag found by Kowal in 1968, but typical of using heterogeneous samples of Type Ia SN as plain standard candles.

At about the same time, Mark Phillips, at CTIO was working in a different direction. Pskovskii (1977 [74], 1984 [75]) had proposed to use the rate of decline after maximum as an additional element to classify SNe, of all types, and suggested that there were correlations between the decline rate and other SN properties. Phillips (1993 [69]) attempted the same kind of study but now with modern data and a much cleaner sample of real Type Ia SNe. He found a strong correlation between the rate of decay after maximum light and both the absolute magnitude of SN Ia at maximum, and the $B-V$ color (an index of the SN temperature). The correlation was strongest in the B band, moderate in V, and minor in R.

This was a very important discovery: Type Ia SN were only approximate standard candles. Yes, there was a sizable intrinsic scattering of the luminosity at maximum light, but it was possible to recognize how far away from the mean value an individual SN was, just from the rate of decay after maximum. As important as this, the same rate of decline was an indicator of the intrinsic color of the SN. This made it possible to correct for the extinction of light caused by foreground interstellar matter in the parent galaxy. Both corrections have a direct impact in improving the character of Type Ia SNe as cosmologically relevant distance estimators. Phillips (1993 [69]) was a major breakthrough.

There were also some hits on the theoretical side, but the advances were limited by computer power. Nomoto, Thielemann & Yokoi (1984 [57]) following up the idea of Nomoto, Sugimoto & Neo (1976 [56]), computed models of carbon deflagration supernovae. One of the main problems they faced was that the deflagration gives raise to Rayleigh-Taylor instabilities and is, hence, a multidimensional process. Forced to treat it in one dimension, they simulated the propagation of the convective carbon deflagration front using a time dependent mixing length theory. They had to assume several numerical parameters to accommodate the hydrodynamics, but followed the nucleosynthesis in detail. One of the models they computed, W7, was a very good match to the observations. W7 was a complete success in terms of explosion energetic, both in the time dependent luminosity output and kinetic energy transferred to the ejecta, but also, for the first time in producing the right amounts of intermediate mass elements at the velocities seen in the early time SN spectra. In particular, synthetic spectra computed using the output produced by the model gave a good fit to the observations, if the outer layers of the SN were mixed during the late stages of the explosion (Branch et al. 1985 [12], Nomoto en al. 1986 [58]). The latter was needed, because the intermediate mass elements in the theoretical model appeared in a very narrow range of velocities, smaller than those observed in real SNe. In spite of this problem, W7 set the standard of comparison for Type Ia SN models for many years to come. Woosley and Weaver (1986 [99]) produced a very influential review paper on SNe in general, centered especially on the physics of explosion and nucleosynthesis, which remains a relevant reference even today.

In addition to the location of the intermediate mass elements, the deflagration models had other important shortcoming: They tended to overproduce ^{54}Fe and other neutron rich Fe peak isotopes, because the burning matter remains at very high temperatures and densities for too long. Also, upon further study, the need to mix the outer layers to take intermediate mass elements to higher velocities became increasingly more difficult to justify (Sutherland and Wheeler 1984). A plain deflagration explosion model, in addition, seemed too constrained to explain the heterogeneity of Type Ia SNe that was starting to appear with improved observations. Khokhlov (1991a [46], 1991b [47]) focused on these problems, and decided to test whether reality could be more complex. He realized that the subsonic nature and related time scales of the deflagration were long enough to change the background where thermonuclear burning was taking place, and, hence, a detonation was likely after a period of deflagration. He named the mechanism delayed detonation and showed that a white dwarfs exploding like that could match the observations better than the plain deflagration of W7. Later study proved that the combination of processes had another virtue: varying the critical density at which the deflagration turned into a detonation (an unknown external parameter in Khokhlov approach), different kinds of Type Ia SNe could be produced matching the diversity of normal, sub-luminous, and luminous events, giving rise to a theoretical interpretation of the Phillips (1993) relation (Höfflich, Khokhlov & Wheeler, 1993 [42]).

Looking with hindsight, by the end of 1993 the field was poised for a breakthrough. Type I SNe had been cleaned so as to clearly isolate the Type Ia events. Theory had advanced enough to solidify

the hypothesis that Ia SNe originate on the thermonuclear combustion of a Chandrasekhar mass white dwarf. This gave strong support to the concept that Ia SNe ought to be very uniform. Reality showed that, nevertheless, there were sources of heterogeneity, but empirical study prompted to a feasible path at calibrating the differences. In addition, theory parametrization of the unknown critical density for a deflagration to turn into a detonation, allowed for a qualitative understanding of these differences, bringing some piece of mind to observational cosmologists who were using the empirical relations without fully grasping its meaning. Solid state detectors were large enough to allow astronomers do relative photometry between extragalactic SN and foreground galactic stars in the field. This made it possible to do extensive follow up even in non photometric conditions.

Finally, the incipient and fast improving internet was a development, not fully recognized at the time, which resulted critical. It allowed astronomers at different observatories and research centers in the world to share results, images and information in real time, making it possible to pool human resources, and observational and computing facilities. This was possible with modest budgets and widespread and heterogeneous funding sources. Hence, advancing a large program of observation and analysis of distant SNe was no longer restricted to large and rich research groups or labs, which could count with all the facilities needed under a centralized management, but became within reach of more horizontal arrays of researchers who could coordinate facilities scattered around the world to work on the same project.

3 A magic SN Cosmology lustrum: 1994-1998

Some groups of SN studies recognized the breakthrough implied by the Phillips (1993 [69]) result. Important as it was, it was based on a small and heterogeneous sample of SN observed with different telescopes and instrumental sets. It was critical then to increase the SN sample and, ideally, observe them under more uniform conditions. Some surveys were started with the specific goal of calibrating the relations of rate of decline versus magnitude, and color, at maximum light. Among the first were the Calán/Tololo Survey (Hamuy et al. 1993 [35]), where M. Phillips himself was a key player, and the long standing SN study program by B. Kirshner, students and post-docs at the Harvard-Smithsonian Center for Astrophysics (CFA). Also, several methods to accomplish the calibration were developed. Among them, the Δm_{15} (Phillips 1993 [69], Hamuy et al. 1996 [37], Phillips et al. 1999 [71]), and the Multicolor Light Curve Shape method, by Riess, Press and Kirshner (1995 [81], 1996 [82]) at Harvard-CFA.

As mentioned earlier, the SCP had started to be effective at finding distant SNe. By 1994 they reported six SNe discovered in approximately six nights of a program combining the 2.5m Isaac Newton and the 4.0m KPNO telescopes (Perlmutter et al. 1994 [63]). In 1995 (Goobar & Perlmutter, 1995 [33]), the group presented a detailed account of an elegant method to use the luminosity distances measured towards SNe to constrain the cosmological parameters Ω_M and Ω_Λ. The method builds contour probability levels comparing cosmological model predictions with an observed Hubble diagram of nearby and distant Type Ia SNe. The paper became a model for the methodology, and most subsequent analysis built up from this study. It showed, however, that the group was still not giving serious quantitative consideration to the "Phillips effect," and to the reddening correction of individual SNe.

It is my personal view, that it was in part the perception of these weaknesses in the analysis planned by the SCP that motivated, by the end of 1994, Brian Schmidt, a postdoctoral fellow at the at Harvard-Smithsonian CFA, and Nick Suntzeff, a senior staff astronomer at Cerro Tololo Inter-American Observatory, to launch an independent enterprise to build a sizable sample of distant

SNe. The group was called "High Z SN Search Team", to emphasize a horizontal constituency. The main goal of this new group was to observe the SNe in a way that allowed a sine qua non spectroscopic classification, a careful understanding of the interstellar extinction in front of each individual SN, a precise measurement of the light curve shape, and a minimization of the systematic effects associated with K-corrections. The first telescope runs of this team were in March 1995, when they discovered SN 1995K at z=0.478, the record holder of the time (Phillips et al. 1995 [68]). I joined the group in September 1995, when moved to CTIO as a postdoctoral fellow.

The years 1996-1997 were thrilling. The SCP and the HZ Team routinely presented proposals to discover and follow up SNe from different telescopes around the globe, including the Hubble Space Telescope. The SCP managed to build what they thought was a small but trustable sample by mid 1997, when they published the paper Measurement of the Cosmological Parameters Omega and Lambda from the First Seven Supernovae at z $\gtrsim$ 0.35 (Perlmutter et al. 1997 [65]). In this paper they present the elegant Stretch Factor Parametrization, and independent way of calibrating the light curve width versus luminosity effect found by Phillips in 1993. They, however, could not correct for extinction the individual events and some of the SN did not have a clear spectroscopic classification as "Type Ia". The results pointed towards a high density universe, where a cosmological constant was excluded with high statistical significance ($\Omega_\Lambda < 0.51$ at the 95% confidence level for a spatially flat universe).

By 1998 the results of the HZ Team started to appear (Garnavich et al. 1998a [30]; Riess et al. 1998 [78]; Schmidt et al. 1998 [84]; Garnavich et al. 1998b [31]) and the times went from thrilling to hectic. The first paper was based on a set of four SNe, three of them observed with HST, and was also presented at the January, 1998, meeting of the American Astronomical Society (Garnavich et al. 1997 [29]). It arrived squarely at the opposite conclusion that Perlmutter et al. (1997 [65]) did: The SN distances were more consistent with a low density Universe that will expand forever. Also by the end of 1997, the SCP was starting to change its view regarding the high density Universe. A single SN well observed with HST and added to the sample of seven SNe published earlier had changed the confidence levels enough to turn their claim into "results [that] are preliminary evidence for a relatively low-mass-density universe" (Perlmutter et al. 1998 [66]). Schmidt et al. (1998 [84]) describes the strategy and analysis method of the HZ Team, and applies them, as a test, to SN 1995K. Riess et al. (1998 [78]) brought the field to a climax presenting the analysis of the first 16 HZ Team distant SNe, with the unavoidable conclusion that the expansion of the Universe is accelerating. A paper submitted in September 1998 by the SCP presented the results based on 42 SNe, and reached essentially the same conclusion (Perlmutter et al. 1999 [67]), confirming finding of an acceleration. The final work of the year by the HZ Team (Garnavich et al. 1998b [31]) set the first constraints on the parameter w in the equation of state of the Universe, based on SN luminosity distances.

The discovery of the Cosmic Acceleration made a big impact. It was selected as the "Science Breakthrough of the Year" for 1998 by Science Magazine (Glanz 1998 [32]). It later meant an impressive sequence of prizes for Perlmutter, Riess and Schmidt, and, in the case of the 2007 Gruber Cosmology Prize, also an explicit acknowledgement for the teams. The string of recognitions culminated in October 2011, with the award of the Nobel Prize in Physics to Perlmutter, Riess and Schmidt.

4 Cosmic Acceleration: The SN result

4.1 "Theory"

Given in the context of a meeting that dealt mostly with modifications of General Relativity that could help to explain perplexing observational results, the following paragraphs seem children's play. However, General Relativity is the framework against which the observations of distant SNe have been compared and the parameters that came out of that comparison have been the source of our puzzlement. So, let us become children again, go back to run-of-the-mill General Relativity and assume, in addition, that the Universe is isotropic and homogeneous. With these hypothesis, the luminosity distance between us and an object located at redshift z can be written (Carroll, Press and Turner, 1992 [18])

$$D_L = \frac{c(1+z)}{H_0\sqrt{|\Omega_k|}}\ \mathrm{sinn}\left\{\sqrt{|\Omega_k|}\int_0^z \left[(1+z)^2(1+\Omega_M z) - z(2+z)\Omega_\Lambda\right]^{-1/2} dz'\right\}, \tag{1}$$

where H_0 is the Hubble constant, Ω_M, Ω_Λ, and Ω_k are the cosmological density parameters for gravitating matter (including baryonic and dark matter), Cosmological Constant, and curvature, respectively, measured in terms of the critical density, and

$$\mathrm{sinn}(x) = \begin{cases} \sinh x & \text{for} \quad \Omega_k > 0 \\ x & \text{for} \quad \Omega_k = 0 \\ \sin x & \text{for} \quad \Omega_k < 0 \end{cases} \tag{2}$$

It is standard usage in astrophysics to express the distances as *Distance Modulus*. This is the difference between the apparent and absolute magnitudes of a source, related with distance by

$$m - M = 5\log\frac{D_L}{10}, \tag{3}$$

where D_L should be given in parsecs. An advantage of the distance modulus is that it is measured directly in magnitudes, so differences in $m - M$ can be directly compared with uncertainties in astrophysical observations and calibrations.

A rapid analysis of how D_L varies with redshift helps to explain the opportunity that the HZ Team founders perceived after the Phillips (1993 [69]) result. Figure 1 shows D_L for some of the different values of the cosmological parameters Ω that were seriously considered in the early nineties, and, in addition, the values of the now called *Concordance* Universe. As became usual in the field, the upper panel displays the run of D_L for different models of the universe, and the lower panel the difference between a given universe and a reference one. Typically the reference universe is the empty, or coasting, universe. It is easy to see in the lower panel of the figure that the difference between models at redshifts of $z \sim 0.5$ is several tens of magnitude. So, with a dispersion of a ~ 0.15 mag per SN, a few tens of SNe at this critical redshift would provide an observation precise to a few hundredth of a magnitude. This point will permit to discriminate between different cosmological models. On the other hand, technology had improved enough by $\sim$1994 that discovering and observing Type Ia SNe at $z \sim 0.5$ was no longer the heroic enterprise that Nørgaard-Nielsen et al. (1988 [59]) had undertaken. The CCDs were larger, more sensitive and, especially, did no have fringing in the red. The latter allowed for a clean image subtraction at the rest-frame blue pass-band at $z \sim 0.5$. This allowed for precise photometry at the pass-band where the light curve shape versus luminosity relation was most sensitive. Finally, the National Optical

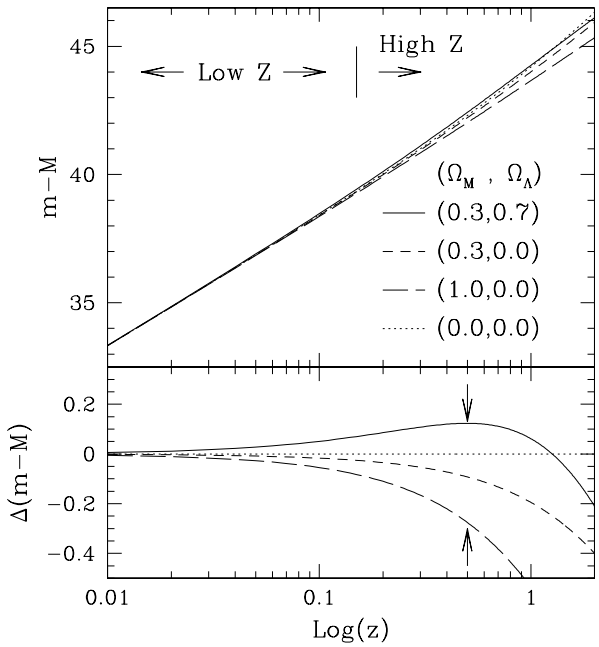


Figure 1: Hubble Diagram of Luminosity distances for various models of the Universe, parametrized by the values of the cosmological density parameters Ω_M, Ω_Λ, under the assumption that the curvature is zero. The upper panel displays the run of D_L and the lower the difference between a given universe and the empty one. In the upper panel the regions considered throughout this paper as low and high redshift are indicated. The vertical arrows in the lower panel mark the position of the critical $z = 0.5$ redshift.

Astronomical Observatory IRAF group had started to develop image matching and subtraction software by the late eighties, and by the early to mid-nineties working moduli became available as prototype IRAF tasks (Phillips and Davies 1995 [68]).

On the other hand, although there is room to discriminate, it is also clear from the figure that systematic effects that change the individual SN results by tenths of a magnitude should not be missed, especially if those effects make you prone to Malmquist bias. Phillips (1993 [69]), Hamuy (1994 [36], 1996 [37]), and Riess, Press and Kirshner (1995 [81]) had shown that Type Ia SN considered normal by any standard had differences in maximum brightness of that order. A magnitude limited search for distant SNe would preferentially discover bright events at large distances and more average SNe nearby. The same would happen if corrections by foreground reddening, which amounts typically to few tens of a magnitude, were not considered. Searches will tend to discover less extincted SNe at large distances and more average extincted SNe nearby. So, if SNe were treated as standard candles and corrections by light curve shape and foreground extinction were not applied, the final comparison would be made between populations of different intrinsic brightness at different redshifts, and the resulting cosmology would be biased.

Finally, equation 1, 3 and Figure 1 allow us to see the key difference between the experiment of measuring the rate of expansion, H_0, and estimating Ω_M and Ω_Λ. Measuring H_0 requires calibrating very precisely the absolute magnitude of, in this case, Type Ia. This is, requires precise knowledge of M. In turn, this calls for accurate measurement of the distance, in physical units, to many galaxies that have hosted Type Ia SNe. Estimating the Ω parameters, on the other hand, requires just a relative measurement. Variations of M and H_0, constants that could be packed together in equation 3, will change the vertical scale of Figure 1, but will not change the relative shape of the different curves. This means that, given the appropriate combinations of telescopes, detectors and computers, estimating H_0 is a more difficult experiment than estimating the Ω parameters.

4.2 Observations: High precision, high redshift Hubble Diagrams

By the end of 1997, the group had discovered more than 20 transients and 14 of them had been positively identified as Type Ia SNe with spectra taken from the Keck Telescope, the Multiple Mirror Telescope, or the European Southern Observatory 3.6m. This set, which included the four SNe that were the basis for Garnavich et al. (1998a [30]), was augmented with SNe from the low redshift sample gathered earlier at the Harvard-CFA (Riess 1996 [77]; Riess et al. 1999 [79]), and by some from the Calán–Tololo Survey (Hamuy et al. 1966 [37]). Between 27 and 34 SNe from these samples were used, depending on the criteria defined to constrain the quality of the light curves. This small high redshift SN set was enough to clearly detect the Cosmic Acceleration (Riess et al. 1998 [78]). The result was soon confirmed by the long awaited SCP full sample paper in 1999 (Perlmutter et al. 1999 [67]), and by many subsequent papers from both teams.

What I will show here are the inferences based on the combined sample of distant SNe that the HZ Team collected throughout its existence (presented in the previously mentioned papers and in Tonry et al. 2003 [89], Barris et al. 2004 [10], and Clocchiatti et al. 2006 [19]), and an expanded sample, including very distant SNe, built by Riess et al. 2004 [80]. Figures 2 and 3 show, 75 nearby and 45 distant Type Ia SNe, and 78 nearby and 109 distant SNe, respectively. The SNe are color coded according to distance and quality of the light curves defined as in Riess et al ([80]. The observed distance moduli are compared with the theoretical predictions for the same models of the universe shown in Figure 1.

The simplest way to interpret the observed SN luminosity distances in empirical terms is to

note that, at a redshift of $z \sim 0.5$ the points tend to be above the theoretical line corresponding to the empty universe. The distance moduli $m - M$ observed, in average, tend to be larger than those predicted, even for a universe that will never decelerate due to its self gravitational attraction. So, SN are further away than expected according with the redshift of their parent galaxies. This observed *excess distance* is interpreted by the models as the effect of an acceleration that has pushed the SNe farther away between the time of their explosions at $z \sim 0.5$ and the present time at $z = 0$.

Also, it is important to note that, even with the spread of the points, the tendency of the SNe is to appear scattered around the empty universe at low redshift (blue points), then preferentially above the empty universe (red points at $z \sim 0.45$ and then preferentially below the empty universe at redshifts $z \sim 1$. This behavior reinforces the strong signal of acceleration, and nullifies explanations based on systematic effects that will make distant SNe to appear progressively more distant. This applies to the "gray" dust hypothesis (Aguirre 1999a [1], 1999b [2]), but also to most systematic effects, since they will typically show a monotonous trend with redshift.

4.3 The best fitting cosmology

In the spirit of Gobbar et al. (1995 [33]) it has become usual to asses the significance of the data for cosmology by using the Hubble Diagrams to compute probability contours for arbitrary sets of cosmological parameters. The particular process I will use is described in Riess et al. (1998). Basically, a set of Ω parameters, and an H_0 are assumed and a, "observed minus expected from theory" χ^2 value is computed using equation 1. The probability of the resulting χ^2 is obtained from the χ^2 distribution and hipper-volume of χ^2 density computed. The volume is then dimensionally reduced by integrating over the "nuisance" parameters, of which H_0 is an example in this case. Eventually, the volume can be projected in two dimensions like in the contour plots shown in Figures 4 and 5.

At first sight, either contour plot shows that the Hubble Diagrams alone do not impose a severe constraints on Ω_Λ or Ω_M individually, but they constrain the difference $\Omega_M - \Omega_\Lambda$ rather tightly. Even though, either contour level indicates with very high confidence that the Universe has $\Omega_\Lambda > 0$. The "excess distance" described in graphical terms by the Hubble Diagrams is quantitatively interpreted, in terms of simple General Relativity, by fitting a significant Ω_Λ.

The observational result is more compelling if the confidence contours from the Hubble Diagrams are combined with those of an independent experiment to constrain either Ω_Λ or Ω_M. In Figure 6, I show the outcome of this exercise assuming as a prior $\Omega_M H_0/100 = 0.20 \pm 0.02$ as found by the Two Degree Field (2dF) Redshift Survey (Percival et al. 2001 [61]). Just this two experiments point strongly to the parameters $\Omega_M \sim 0.3$ and $\Omega_\Lambda \sim 0.7$ which, together with $\Omega_K = 0$ and your favorite prescription for the baryonic matter density, have been lately known as the "Concordance Universe."

5 Is it Ω_Λ or a more general *Dark Energy*?

5.1 The "Second Generation" Surveys

The success at detecting and measuring the Cosmological Constant using high precision Hubble Diagrams of distant Type Ia SNe prompted astrophysicists to go a step further. SNe appear now to be so precise as distance indicators that the goal of testing whether Ω_Λ is constant, or not, in time appears to be possible. On the theoretical side, many pressing questions are risen by a constant

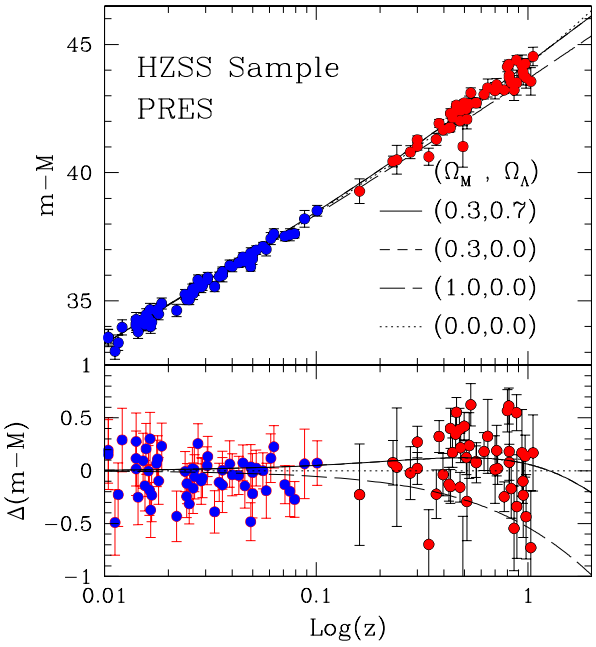


Figure 2: Hubble Diagram of the 45 distant SNe collected by the High Z Supernova Search Team between 1995 and 2002 (red points), together with 58 SNe from the local sample (blue points). Distances for this plot were computed according with the PRES method (Prieto, Rest and Suntzeff 2003 [72]), an elegant generalization of the Δm_{15} method. The theoretical luminosity distances for universes with different density parameters Ω_M, Ω_Λ, are also shown, as in Figure 1.

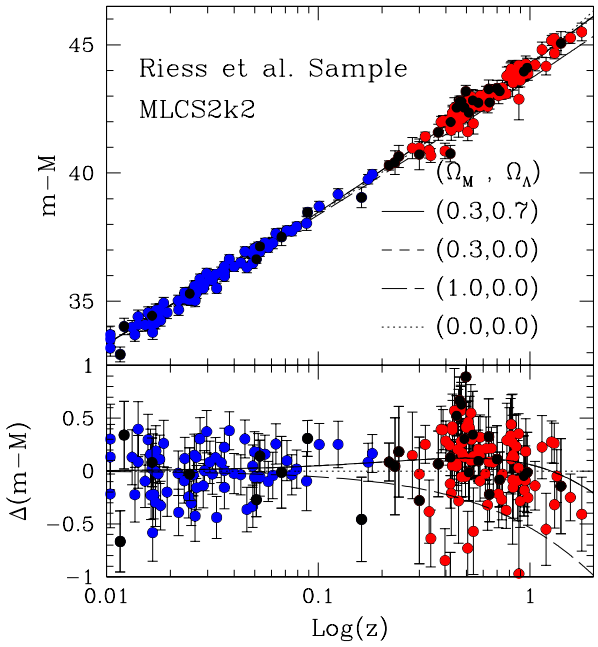


Figure 3: Same as Figure 2, but for the 78 nearby and 109 distant SNe of the sample of Riess at al. 2004 [80] In this case the distance were computed by the method MLCS2k2 (Jha et al. 2007 [44]). Black symbols indicate SNe which do not qualify as "gold" according with the criteria of Riess et al. 2004 [80].

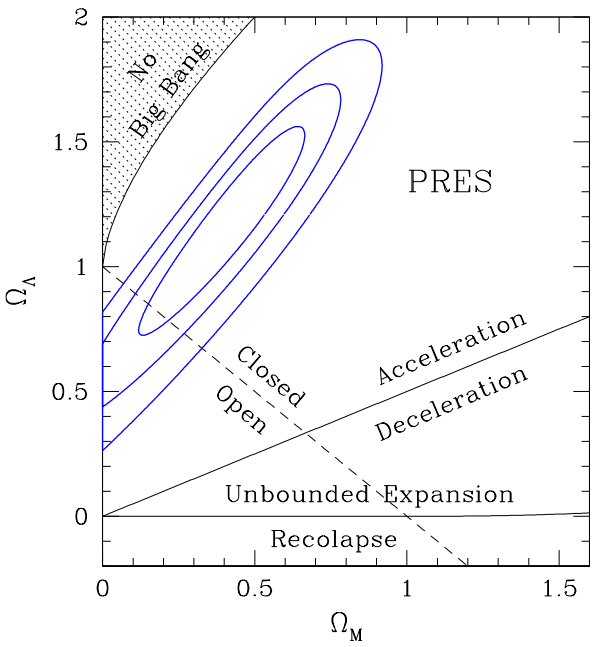


Figure 4: Joint probability contours for the parameters Ω_Λ and Ω_M that best fit the Hubble diagram of Figure 2 (distances calibrated using the PRES method). From larger to smaller, the drawn contours correspond to 99.5%, 97%, and 68% confidence, respectively.

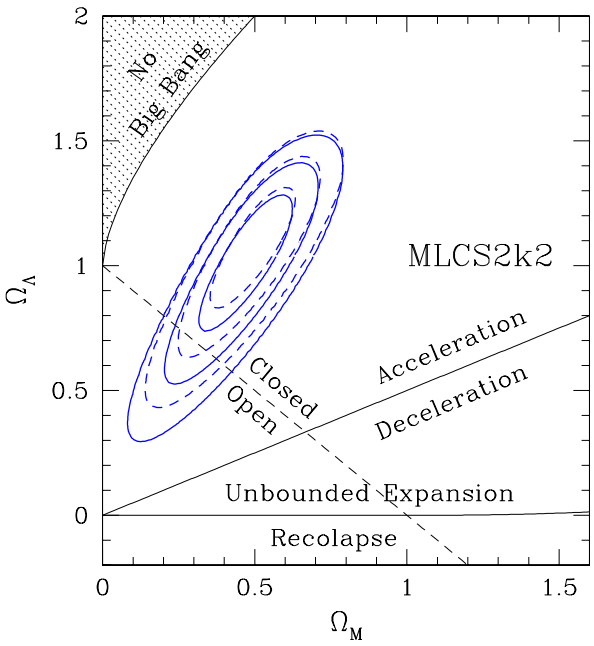


Figure 5: Joint probability contours for the parameters Ω_Λ and Ω_M that best fit the Hubble diagram of Figure 3 (distances calibrated using the MLCS2k2 method). The contours drawn with solid line include only those SNe that qualify as "gold" according with Riess et al. (2004 [80]). The dashed contours include "gold" and "silver" SNe. Again, from larger to smaller, the contours drawn correspond to 99.5%, 97%, and 68% confidence, respectively.

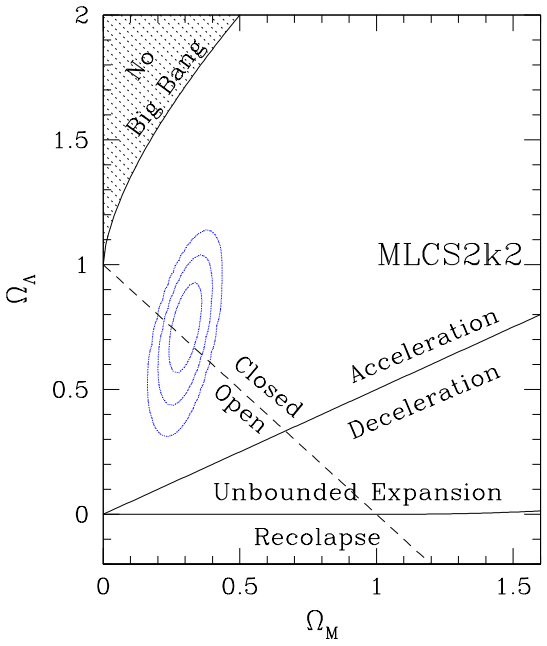


Figure 6: Joint probability contours for the parameters Ω_Λ and Ω_M of the "gold" sample of Riess et al. (2004 [80]) (see previous figure), but now computed assuming the 2dF prior in Ω_M. As before, from larger to smaller, the contours drawn correspond to 99.5%, 97%, and 68% confidence, respectively.

vacuum energy. The more compelling appear to be the problems of scale and timing. Carroll (2001 [17]) presents a detailed review. One of the possible solutions to the problems posed by a Cosmological Constant is that the "Dark Energy" is not just a constant, but some other kind of field.

From the point of view of an observational astrophysicist, the experiment starts by parametrizing the luminosity distances in terms of this unknown energy density. It is natural to write the relation between pressure and density using the fairly general form of the equation of state that applies to fluids

$$P = w\rho, \tag{4}$$

where P is the pressure, ρ the density, and w a constant. From this form and the energy-momentum equation it is found that ρ evolves with the scale factor a of the Universe as

$$\rho \propto a^{-n}, \tag{5}$$

where

$$n = 3(1 + w). \tag{6}$$

In this parametrization, gravitational matter ($n = 3$) requires $w = 0$, radiation ($n = 4$) implies $w = 1/3$, and a Cosmological Constant ($n = 0$), will mean $w = -1$.

Introducing as before the cosmological density parameters by dividing the density of the different components into the critical density, the equation 1 can be expressed as

$$D_L = \frac{c(1+z)}{H_0\sqrt{|\Omega_k|}}\ \mathrm{sinn}\left\{\sqrt{|\Omega_k|}\int_0^z \left[\sum_i \Omega_i(1+z')^{n_i} + \Omega_k(1+z')^2\right]^{-1/2} dz'\right\}, \tag{7}$$

where the sub index i denotes the different density components and the sub index k, as before, is reserved for the curvature.

Equation 7 can be used in the same way as equation 1 to set up a minimization problem and simultaneously constrain Ω_M and w, instead of Ω_M and Ω_Λ. I will use the sample of SNe presented before to illustrate the nature of the problem. Figure 7 show the confidence contours computed with the "gold" sample of Riess et al. (2004 [80]) with and without the additional constrain of an independent measurement of Ω_M. A quick inspection of the figure reveals that measuring w is hard. When applied to the problem of Ω_Λ, the sample produces the results of Figures 5 and 6, which appear as a strong signal for the Cosmological Constant with $\Omega_\Lambda \sim 0.7$. It is tough, however, to reach any meaningful conclusion about w from the contours of Figure 7: the one sigma uncertainty is $\sim$60%. This kind of constrain is consistent with a very wide range of Dark Energy models.

The brute force approach to improve upon this result is to increase the number of good quality SN in the sample, from many tens to a few hundred. By the early 21st Century, two new collaborations appeared, with ex-members of both the SCP and the High Z SN Search Team rearranged among them, and organized new projects to try to answer the more complex question: Is the Cosmic Acceleration caused by a Cosmological Constant or not? The newer projects were named SN Legacy Project (Astier et al. 2006 [5]) and ESSENCE (Miknaitis et al. 2007 [50]). Both collaborations have been working hard to understand the systematic biases, which, with the much increased number of SNe, will be the dominant source of uncertainty. Interestingly, as the distant SN sample reaches the many hundreds to the thousands, the nearby sample of around a hundred SNe becomes one of the sources of systematic uncertainty. Fortunately, too, some collaborations have also appeared to enlarge it (Hamuy et al. 2006 [38], Rau et al. 2009 [76]).

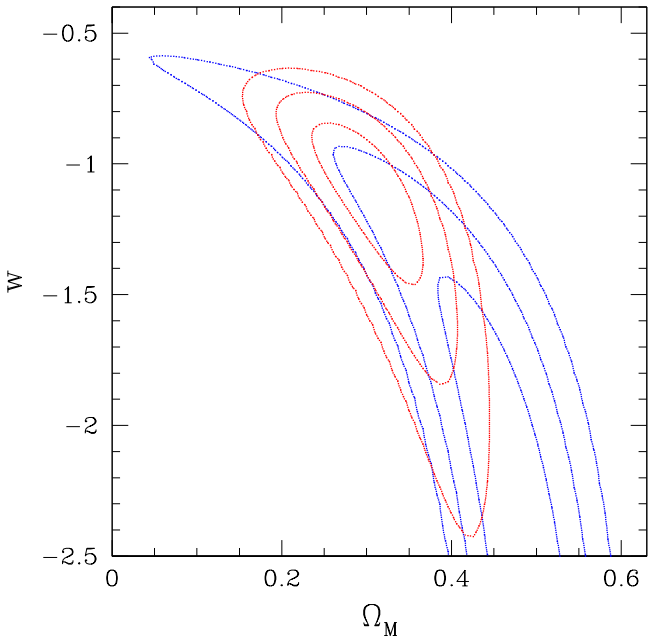


Figure 7: Joint probability contours for the parameters w and Ω_M for the "gold" sample of Riess et al. (2004). The blue contours which extend below $w = -2.5$ are computed only with the SN sample. The red ones were computed assuming the 2dF prior in Ω_M. As in all cases before, from larger to smaller, the contours drawn correspond to 99.5%, 97%, and 68% confidence, respectively.

The early results, presented in several additional papers (Wood-Vasey et al. 2007 [98], Guy et al. 2010 [34], Sullivan et al. 2011 [86]) have taken the one sigma uncertainty down to $\sim$7% and have been persintetly concluding that w is fully consistent with -1, this is, the plain Cosmological Constant.

5.2 Conclusion

Shall we call the problem solved and convince ourselves that the Cosmic Acceleration is, really, caused by a Cosmological Constant? My personal view is that the case for it is strong, but still not certain. As shown by the analysis of Miknaitis et al. (2007 [50]), we have not controlled yet all the possible systematic uncertainties down to the very demanding limits required by this experiment.

A few years ago we confronted a similar situation when pondering just the existence of Ω_Λ. The theoretical expectations for it was so enormous that the very low observational limits existing before 1998 were taken as a strong indication that $\Omega_\Lambda = 0$. It was much easier to contrive mechanisms that will completely cancel it, than imagine ways to make it much smaller than expected, but not zero. Many of us were *convinced* that $\Omega_\Lambda = 0$. Now, the observations indicate that $w \simeq -1$, making it difficult for us to resist the urge to jump at the conclusion that, then, $w = -1$. We need to pay attention to the fact that the cosmic acceleration has fooled us more than once in the past, and that there is still observational room for her to do it again. The task for the observers during the next five to ten years is to develop better instruments, better experiments, and to keep a paranoid eye on the sources of systematic effects.

6 Acknowledgements

All of my work in the extremely exciting, demanding, challenging and competitive realm of SN Cosmology has been done as a collaborator of two fairly large teams. It started in 1995, when I was a postdoctoral fellow at Cerro Tololo Inter-American Observatory in Chile. I was invited by Brian Schmidt to join the High Z Supernova Search Team, a now famous group of about 20 astrophysicists who were trying to measure q_0, the *deceleration* parameter of the Universe. It continued with the ESSENCE project, carried on by an even larger team that grew out of the HZ Team in 2001 and keeps on working to improve the SN constrains on cosmology.

Support for my research has been mainly provided by Chile, through many different grants. The active ones are Iniciativa Científica Milenio P10-064-F (MINECON), Programa de Fondos Basales CATA PFB 06/09 (CONICYT), and FONDAP No. 15010003 (CONICYT).

References

[1] Aguirre, A., *ApJ*, **512**, L19 (1999a).

[2] Aguirre, A., *ApJ*, **525**, 583 (1999b).

[3] Arnett, W.D., *Ph.D. Diss.*, Yale University (1966).

[4] Arnett, W.D., *ApSS* **5**, 180 (1969).

[5] Astier, P., et al., *AA*, **447**, 31 (2006).

[6] Baade, W., *Ap.J.*, **102**, 309 (1945).

[7] Baade, W., Burbidge, G., Hoyle, F., Burbidge, M., Christy, R., Fowler, W., *PASP*, **68**, 296 (1956).

[8] Baade, W. & Zwicky, F., *Proc. N.A.S.*, **20**, 259 (1934).

[9] Baade, W. & Zwicky, F., *Phys. Rev.*, **46**, 76 (1934)

[10] Barris, B., et al., *ApJ*, **602**, 571 (2004).

[11] Borst, L.B., *PhRv*, **78**, 807 (1950).

[12] Branch, D., Dogget, J.B., Nomoto, K. & Thielemann, F.K. *Ap.J.*, **294**, 619 (1985).

[13] Branch, D., Lacy, C., McCall, M., Sutherlnad, P., Uomoto, A., Wheeler, J.C. & Willis, B.J., *Ap.J.*, **270**, 123 (1983).

[14] Branch, D. & Patchett, B., *MNRAS*, **161**, 71 (1973).

[15] Burbidge, E., Burbidge, G. & Fowler, W. *RMP*, **29**, 547 (1956).

[16] Burbidge, G., Hoyle, F., Burbidge, M. & Christy, R. *Phys. Rev.*, **103**, 1145 (1956).

[17] Carroll, S.M., *LRR*, **4**, 1 (1992).

[18] Carroll, S.M., Press, W., Turner, E.L., *ARAA*, **30**, 499 (1992).

[19] Clocchiatti, A., et al. *Ap.J.*, **642**, 1 (2006).

[20] Colgate, S.A., *Ap.J.*, **232**, 404 (1979).

[21] Colgate, S.A., Grasberger, W.H. & White, R.H., *AJ*, **66**, 280 (1961).

[22] Colgate, S.A., Grasberger, W.H. & White, R.H., *Journal of the Phys. Soc. of Japan*, **17**, 157 (1962).

[23] Colgate, S.A. & Mc Kee, C., *Ap.J.*, **157**, 623 (1969).

[24] Colgate, S.A. & White, R.H., *Ap.J.*, **143**, 626 (1966).

[25] Couch, W.J., Perlmutter, S., Newburg, H.J.M., Pennypacker, C., Goldhaber, G., Muller R. & Boyle, B.J., *PASAu*, **9**, 261 (1991).

[26] Elias, J.H., Frogel, J.A. & Hackwell, J.A., *Ap.J.*, **251**, L13 (1981).

[27] Elias, J.H., Matthews, K., Neugebauer, G. & Persson, S.E. *Ap.J.*, **296**, 379 (1985).

[28] Finzi, A., & Wolf, R.A., *Ap.J.*, **150**, 115 (1967).

[29] Garnavich, P., et al., *AAS*, **191**, 850 (1997).

[30] Garnavich, P., et al., *Ap.J.*, **493**, L53 (1998a).

[31] Garnavich, P., et al., *Ap.J.*, **509**, 74 (1998b).

[32] Glanz, J., *Science*, **282**, 2156 (1998).

[33] Goobar, A., & Perlmutter, S., *ApJ*, **450**, 14 (1995).

[34] Guy, S.M., *AA*, **523**, 7 (2010).

[35] Hamuy, M., et al., *A.J.*, **106**, 2392 (1993).

[36] Hamuy, M., et al., *AJ*, **108**, 2226 (1994).

[37] Hamuy, M., et al., *AJ*, **112**, 240 (1996).

[38] Hamuy, M., et al., *PASP*, **118**, 2 (2096).

[39] Hansen, C.J. & Wheeler, J.C., *ApSS*, **3**, 464 (1969).

[40] Hansen, L., Jorgensen, H.E. & Norgaard-Nielsen, H.U., *ESO Messenger*, **47**, 46 (1987).

[41] Harkness R.P., et al., *Ap.J.*, **317**, 355 (1987).

[42] Höfflich, P., Khokhlov, A. & Wheeler, J.C., *Ap.J.*, **444**, 831 (1993).

[43] Hoyle, F. & Fowler, W., *Ap.J.*, **132**, 565 (1960).

[44] Jha, S., Riess, A.G., & Kirshner, R.P., *Ap.J.*, **659**, 122 (2007).

[45] Kare, J.T., Pennypacker, C.R., Muller, R.A., Mast, T.S., Crawford, F.S. & Burns, M.S. *NATO Adv. Study Inst. on SNe*, Cambridge, Jun 28^{th} (1981).

[46] Khokhlov, A., *A&A*, **245**, 114 (1991a).

[47] Khokhlov, A., *A&A*, **245**, L25 (1991b).

[48] Kowal, C.T., *AJ*, **73**, 1021 (1968).

[49] McLaughlin, D., *PASP*, **75**, 133 (1963).

[50] Miknaitis, G., et al., *ApJ*, **666**, 674 (2007).

[51] Minkowsky, R., *PASP*, **53**, 224 (1941).

[52] Minkowski, R., *PASP*, **75**, 505 (1963).

[53] Mustel, E.R., *SvA*, **15**, 1 (1971a)

[54] Mustel, E.R., *SvA*, **15**, 527 (1971b).

[55] Mustel, E.R., *SvA*, **16**, 10 (1972).

[56] Nomoto, K., Sugimoto, D., & Neo, S., *ApSS*, **39**, 37 (1976).

[57] Nomoto, K., Thielemann, F.K. & Yokoy, K., *Ap.J.*, **286**, 644 (1984).

[58] Nomoto, K., Thielemann, F.K., Yokoy, K. & Branch, D., *Ap. & S.S.*, **118**, 305 (1986).

[59] Norgaard-Nielsen, H.U., Hansen, L., Jorgensen, H.E., Aragon Salamanca, A. & Ellis, R.S., *Nature*, **339**, 523 (1989).

[60] Payne-Gaposchkin, C. & Whipple, F., *PNAS*, **26**, 264 (1940).

[61] Percival,W. J., et al., *MNRAS*, **327**, 1297 (2001).

[62] Perlmutter, S., et al., *Particle Astrophysics: Forefront Experimental Issues*, E.B. Norman, ed., Singapore: World Scientific, 1989, 196 (1989).

[63] Perlmutter, S., et al., *BAAS*, **26**, 965 (1994).

[64] Perlmutter, S., et al., *Ap.J.*, **440**, L41 (1995)

[65] Perlmutter, S., et al., *Ap.J.*, **483**, 565 (1997).

[66] Perlmutter, A., et al., *Nature*, **391**, 51 (1998).

[67] Perlmutter, A., et al., *Ap.J.*, **517** 565 (1999).

[68] Phillips, A.C., & Davies, L.E., *ASPC*, **77**, 297 (1995).

[69] Phillips, M.M., *Ap.J.*, **413**, L105 (1993).

[70] Phillips, M.M., *IAU Circ*, 6160 (1995).

[71] Phillips, M.M., Lira, P., Suntzeff, N.B., Schommer, R.A., Hamuy, M., Maza, J., *AJ*, **118**, 1766 (1999).

[72] Prieto, J.L., Rest, A., & Suntzeff, N.B., *ApJ*, **647**, 501 (2006).

[73] Pskovskii, Yu. P., *Sv.A.*, **12**, 750 (1969).

[74] Pskovskii, Yu. P., *Sv.A.*, **21**, 675 (1977).

[75] Pskovskii, Yu. P., *Sv.A.*, **28**, 658 (1984).

[76] Rau, A., et al., *PASP*, **121**, 1334 (2009).

[77] Riess, A.G., *Ph.D. Thesis, Harvard University* (1996).

[78] Riess, A.G., et al., *AJ*, **116**, 100 (1998).

[79] Riess, A.G., et al., *AJ*, **117**, 707 (1999).

[80] Riess, A.G., et al., *ApJ*, **607**, 665 (2004).

[81] Riess, A.G., Press W.H. & Kirshner, R.P. *Ap.J.*, **438**, 17 (1995).

[82] Riess, A.G., Press, W.H. & Kirshner, R.P. *Ap.J.*, **473**, 88 (1996).

[83] Sandage, A., & Tammann, G.A. *ApJ*, **415**, 1 (1993).

[84] Schmidt, B.P., et al., *Ap.J.*, **507**, 46 (1998).

[85] Suess, H.E., Urey, H.C., *RvMP*, **28**, 53 (1956).

[86] Sullivan, M., et al., *ApJ*, **737**, 102, (2011).

[87] Sutherland, P.G. & Wheeler, J.C., *Ap.J.*, **280**, 282 (1984).

[88] Tammann, G.A., in *Scientific Research with the Space Telescope*, ed.M.S. Longair and J.W. Warner, 263-293 (1979).

[89] Tonry, J., et al. *ApJ*, **594**, 1 (2003).

[90] Truran, J.W., *Ph.D. Diss., Yale University* (1966).

[91] Truran, J.W., Arnett, W.D. & Cameron, A.G.W., *Ca.J.Ph.*, **45**, 2315 (1967).

[92] Truran, J.W. & Cameron, A.G.W., *ApSS*, **14**, 179 (1971).

[93] Wellmann, P., *ZA*, **35**, 205 (1995).

[94] Wheeler, J.C. & Hansen, C.J., *ApSS*, **11**, 373 (1971).

[95] Wheeler, J.C. & Harkness, R.P., *Pr.Phys.*, **53**, 1467 (1990).

[96] Wheeler, J.C. & Levreault, R., *Ap.J.*, **294**, L17 (1985).

[97] Whipple, F. & Payne-Gaposchkin, C., *PNAS*, **84**, 1 (1941).

[98] Wood-Vasey W.M., et al. *ApJ*, **666**, 694 (2003).

[99] Woosley, S.E., & Weaver, T.A., *ARAA*, **24**, 205 (1986).

[100] Zwicky, F., *PASP*, **73**, 351 (1961).

[101] Zwicky, F., *Stellar Structure*, Chap. 7, eds. L.H. Haller, D.B. McLaughin, University of Chicago Press, Chicago (1965).

Intrinsically Quantum-Mechanical Gravity and the Cosmological Constant Problem

Philip D. Mannheim

Department of Physics, University of Connecticut, Storrs, CT 06269, USA
philip.mannheim@uconn.edu

Abstract. We propose that gravity be intrinsically quantum-mechanical, so that in the absence of quantum mechanics the geometry of the universe would be Minkowski. We show that in such a situation gravity does not require any independent quantization of its own, with it being quantized simply by virtue of its being coupled to the quantized matter fields that serve as its source. We show that when the gravitational and matter fields possess an underlying conformal symmetry, the gravitational field and fermionic matter-field zero-point fluctuations cancel each other identically. Then, when the fermions acquire mass by a dynamical symmetry breaking procedure that induces a cosmological constant in such conformal theories, the zero-point fluctuations readjust so as to cancel the induced cosmological constant identically. The zero-point vacuum problem and the cosmological constant vacuum problems thus mutually solve each other. We illustrate our ideas in a completely solvable conformal-invariant model, namely two-dimensional quantum Einstein gravity coupled to a Nambu-Jona-Lasinio self-consistent fermion.

Keywords: conformal gravity, quantum gravity, cosmological constant problem
PACS: 04.60.-m, 04.50.Kd, 04.90.+e

1. STATEMENT OF THE PROBLEM

Included in the gravitational sources that are commonly used in astrophysics and cosmology are some intrinsically quantum-mechanical ones such as the electron Pauli degeneracy pressure that stabilizes white dwarf stars and the black-body radiation energy density and pressure that contribute to cosmic expansion. As such, these sources contribute to the matter energy-momentum tensor $T_{\rm M}^{\mu\nu}$, and thus if the Einstein equations $(1/\kappa_4^2)G^{\mu\nu}+T_{\rm M}^{\mu\nu}=0$ are to be treated as operator identities, they would require $G^{\mu\nu}$ to be quantum-mechanical too. But since radiative corrections to quantum Einstein gravity are not renormalizable, by hand one instead posits that the Einstein equations are to be understood as being of the semi-classical form $(1/\kappa_4^2)G_{\rm CL}^{\mu\nu}+\langle Q|T_{\rm M}^{\mu\nu}|Q\rangle=0$, with a classical $G_{\rm CL}^{\mu\nu}$ coupling to a c-number matrix element of $T_{\rm M}^{\mu\nu}$ in the quantum state $|Q\rangle$ of interest. However, since the quantum-mechanical $T_{\rm M}^{\mu\nu}$ involves products of fields at the same point, its matrix elements are not finite, and thus even though gravity couples to energy and not energy difference, in addition one equally by hand subtracts off the divergent zero-point vacuum part, to yield

$$\frac{1}{\kappa_4^2}G_{\rm CL}^{\mu\nu}+\langle Q|T_{\rm M}^{\mu\nu}|Q\rangle-\langle\Omega|T_{\rm M}^{\mu\nu}|\Omega\rangle_{\rm DIV}=0 \tag{1}$$

where $|\Omega\rangle$ is the matter field vacuum. It is in the generic form (1) (as augmented by any classical $T^{\mu\nu}_{\rm CL}$ that might also be present) that applications of standard gravity are conventionally made. Thus, for a generic set of oscillators with Hamiltonian $H = \sum(a^\dagger a + 1/2)\hbar\omega$, in taking one-particle matrix elements in states of the form $a^\dagger|\Omega\rangle$ one keeps the contribution of the $a^\dagger a\hbar\omega$ term and ignores the $\hbar\omega/2$ term. However, even if one does render the vacuum quantity $\langle\Omega|T^{\mu\nu}_{\rm M}|\Omega\rangle - \langle\Omega|T^{\mu\nu}_{\rm M}|\Omega\rangle_{\rm DIV}$ finite this way, the Lorentz invariance of the vacuum state still permits this quantity be of the form $-\Lambda g^{\mu\nu}$, with the theory thus having to possess a cosmological constant Λ associated with the essentially uncontrollable finite part of $\langle\Omega|T^{\mu\nu}_{\rm M}|\Omega\rangle$. Moreover, as a cooling universe goes through a cosmological phase transition, not only is an additional, potentially enormous, contribution to Λ then induced, in addition new zero-point infinities are induced in $\langle Q|T^{\mu\nu}_{\rm M}|Q\rangle$ due to mass generation ($\omega = |\bar{k}|$ being replaced by $\omega = (k^2 + m^2/\hbar^2)^{1/2}$), with a one-time subtraction term $\langle\Omega|T^{\mu\nu}_{\rm M}|\Omega\rangle_{\rm DIV}$ not being able to cancel all infinities or readily control the finite part of $\langle Q|T^{\mu\nu}_{\rm M}|Q\rangle - \langle\Omega|T^{\mu\nu}_{\rm M}|\Omega\rangle_{\rm DIV}$ at temperatures both above and below the transition temperature. The challenge to standard gravity then is to naturally recover (1) starting from a fundamental quantum gravitational theory in a way that would both clarify the nature of the subtraction procedure and naturally lead to the small value for Λ that the theory phenomenologically requires. Since this challenge has yet to be met, in this paper we shall propose an alternate approach, one in which the difficulties associated with (1) are by-passed by not having an equation such as (1) appear at all.

To achieve this specific objective we will need to be able to construct a quantum gravitational theory that is consistent and renormalizable, so that we will then be able to use gravity itself to cancel the matter field zero-point fluctuations. Thus once one has a consistent quantum gravity theory, one then has controllable gravitational zero-point fluctuations that are available to effect the needed cancellations. Generically, if we define the action of the universe to be of the form $I_{\rm UNIV} = I_{\rm GRAV} + I_{\rm M}$, then on defining the functional variation of each one of these terms with respect to the metric to be its associated energy-momentum tensor, stationarity with respect to the metric then yields the condition

$$T^{\mu\nu}_{\rm UNIV} = T^{\mu\nu}_{\rm GRAV} + T^{\mu\nu}_{\rm M} = 0. \tag{2}$$

In theories in which the full gravitational plus matter action is renormalizable, the vanishing of the total $T^{\mu\nu}_{\rm UNIV}$ of the universe would survive radiative corrections and serve as an operator identity. It would thus hold in any state, and immediately lead to vacuum cancellation in the form $\langle\Omega|T^{\mu\nu}_{\rm GRAV}|\Omega\rangle + \langle\Omega|T^{\mu\nu}_{\rm M}|\Omega\rangle = 0$, with the zero-point contributions of the gravitational and matter fields identically canceling each other, and with each field serving to regulate the other's divergences. Moreover, in the event of a change in vacuum to some spontaneously broken vacuum $|S\rangle$, the stationarity condition would continue to hold in the form $\langle S|T^{\mu\nu}_{\rm GRAV}|S\rangle + \langle S|T^{\mu\nu}_{\rm M}|S\rangle = 0$, and thus while there would now be both mass generation and cosmological constant generation, all the various zero-point contributions would have to readjust in precisely the manner needed in order to continue to maintain the overall cancellation. The solution to the cosmological constant problem then is to treat the cosmological constant term in conjunction with the zero-point fluctuations, and in this paper we will show how this explicitly happens in a very simple solvable model. To contrast this approach with one based on (1), we see that in (1) the

zero-point fluctuations are discarded before one even begins to tackle the cosmological constant problem, while in (2) the zero-point fluctuations play the central role. Finally, if (2) does hold as an operator identity, then its matrix elements in states with an indefinite number of gravitational quanta will lead to a macroscopic c-number gravitational theory that will serve as the associated classical gravity theory, in complete accord with the way one is able to transit from quantum to classical electrodynamics by taking matrix elements in states with an indefinite number of photons.

In analyses based on (1), we note that already in flat spacetime the quantity $\langle\Omega|T_{\rm M}^{\mu\nu}|\Omega\rangle$ will possess zero-point contributions while the quantity $\langle\Omega|T_{\rm M}^{\mu\nu}|\Omega\rangle - \langle\Omega|T_{\rm M}^{\mu\nu}|\Omega\rangle_{\rm DIV}$ will possess a $-\Lambda\eta^{\mu\nu}$ term where $\eta^{\mu\nu}$ is the Minkowski metric. With such vacuum terms occurring even in the absence of gravity, on expanding $G_{\rm CL}^{\mu\nu}$ as a power series in Newton's constant κ_4^2, we see that gravity can only respond to these vacuum contributions but not control them. Indeed, it is precisely because of issues like this that the cosmological constant problem has proven to be so hard to solve, with it being very difficult for gravity to solve a problem that it is not responsible for. Moreover, if fundamental scalar Higgs fields exist, then $T_{\rm M}^{\mu\nu}$ will even contain a classical piece as well, to provide yet another term over which gravity would have no control. In order to give gravity control of the problems that afflict it, we thus propose to put gravity on an equal footing with matter by expanding the metric not as a power series in the gravitational coupling constant but as a power series in Planck's constant instead. Additionally, we propose that in the absence of quantum mechanics there would be no curvature at all with all the mass and length scales needed to characterize spacetime curvature being intrinsically quantum-mechanical, so that in the absence of quantum mechanics the geometry would be Minkowski. Then, with curvature only occuring in the presence of $\hbar$ there can be no classical contributions to $T_{\rm M}^{\mu\nu}$, with any mass-generating symmetry breaking (viz. the mechanism that generates mass and length scales in the first place) needing to be effected via dynamical fermion condensates rather than by fundamental Higgs fields. Thus we propose that $T_{\rm M}^{\mu\nu}$ and $T_{\rm GRAV}^{\mu\nu}$ both be intrinsically quantum-mechanical with neither containing any intrinsic classical contributions whatsoever. As we will see, this will lead us to a natural resolution of the vacuum energy problem, and as a bonus we will find that we do not need to quantize gravity independently. Rather, once $T_{\rm M}^{\mu\nu}$ is quantized, $T_{\rm GRAV}^{\mu\nu}$ will be quantized simply by virtue of its being coupled to $T_{\rm M}^{\mu\nu}$ in (2). Moreover, with there being no intrinsic classical gravity, one no longer needs to address the issue of how quantization might affect some given classical gravitational configuration.

As regards possible theories of quantum gravity that we could consider, we note that apart from string theory (which we do not consider here), there are three other known theories with controllable radiative corrections, namely four-dimensional conformal gravity with Weyl action $I_{\rm W} = -\alpha_g \int d^4x(-g)^{1/2} C_{\lambda\mu\nu\kappa}C^{\lambda\mu\nu\kappa}$ (see e.g. [1]), its conformal supergravity extension, and Einstein-Hilbert gravity with action $I_{\rm GRAV} = -(1/2\kappa_2^2)\int d^2x(-g)^{1/2}R^{\alpha}{}_{\alpha}$ in two dimensions. Unlike string theory, all of these other three theories are conventional local field theories, and each one of them is locally conformal invariant. Moreover, it will be this very conformal symmetry that will play a central role in the following, with the vanishing of the trace of the matter field energy-momentum tensor obliging its various vacuum contributions to have to mutually cancel

each other identically. Moreover, this same underlying conformal invariance forbids the presence of any possible fundamental cosmological constant term at the level of the input action or the presence in it of any fundamental double-well Higgs potential with its tachyonic mass, making conformal invariance at the level of the action an ideal starting point to attack the cosmological constant problem. Now as regards four-dimensional conformal gravity, it had been thought that because its equations of motion are fourth order, a quantum gravity theory based on it would not be unitary. However, on explicitly constructing the relevant Hilbert space, the theory was found [2, 3, 4, 5] to be free of both negative norm and negative energy states, and it can thus be recognized as a viable quantum gravity theory. However, for the purposes of this paper we will only study the much simpler two-dimensional theory as it is serves to illustrate all of the key ideas presented here, and report the analog discussion of the fourth-order theory itself in [4, 5, 6].

2. ZERO-POINT ENERGY DENSITY AND ZERO-POINT PRESSURE

To illustrate the nature and tensor structure of the vacuum issues that are involved, it is convenient to first look at the vacuum expectation value of the energy-momentum tensor $T_{\rm M}^{\mu\nu} = i\hbar\bar{\psi}\gamma^\mu\partial^\nu\psi$ of a free fermion of mass m in flat, four-dimensional spacetime. With $k^\mu = ((k^2 + m^2/\hbar^2)^{1/2}, \bar{k})$ it evaluates to

$$\langle\Omega|T_{\rm M}^{\mu\nu}|\Omega\rangle = -\frac{2\hbar}{(2\pi)^3}\int_{-\infty}^{\infty} d^3k \frac{k^\mu k^\nu}{k^0}. \tag{3}$$

In (3) we recognize a matter-field zero-point energy density $\rho_{\rm M} = \langle\Omega|T_{\rm M}^{00}|\Omega\rangle$ and a matter-field zero-point pressure $p_{\rm M} = \langle\Omega|T_{\rm M}^{11}|\Omega\rangle = \langle\Omega|T_{\rm M}^{22}|\Omega\rangle = \langle\Omega|T_{\rm M}^{33}|\Omega\rangle$, with each of these quantities being divergent. With its $k^\mu k^\nu$ structure, $\langle\Omega|T_{\rm M}^{\mu\nu}|\Omega\rangle$ has the generic form of a perfect fluid with a timelike fluid velocity vector $U^\mu = (1,0,0,0)$, viz.

$$\langle\Omega|T_{\rm M}^{\mu\nu}|\Omega\rangle = (\rho_{\rm M} + p_{\rm M})U^\mu U^\nu + p\eta^{\mu\nu}, \quad \eta_{\mu\nu}\langle\Omega|T_{\rm M}^{\mu\nu}|\Omega\rangle = 3p_{\rm M} - \rho_{\rm M}. \tag{4}$$

The presence of the timelike fluid 4-vector in (4) is due to the fact that the integration in (3) is over on-shell fermion modes, to thus be a three-dimensional integration and not a four-dimensional one, with the time and space components of k^μ being treated differently.

Even though (3) involves terms that are infinite and thus not well-defined, we note the generic perfect fluid form given in (4) can be established by integrating over the direction of the 3-momentum vector $\bar{k}$ alone, an integration that is completely finite. The perfect fluid form for (3) can thus be established prior to the subsequent divergent integration over the magnitude of the momentum, with this latter integration itself not then reducing the two infinities contained in $\langle\Omega|T_{\rm M}^{\mu\nu}|\Omega\rangle$ to the single one that would be contained in the pure $-\Lambda\eta^{\mu\nu}$ form associated with a cosmological constant. Even though (3) is not well-defined, for the purposes of this paper the perfect fluid form given in (4) is a very convenient way of summarizing the infinities in (3) that need to be cancelled.

As such, the vacuum value given in (3) does not have the form of a cosmological constant term, and indeed it could not since when the fermion is massless the trace of $\langle\Omega|T^{\mu\nu}_{\rm M}|\Omega\rangle$ would be zero while the trace of $-\Lambda\eta^{\mu\nu}$ is given by -4Λ. Moreover, since gravity couples to the full $T^{\mu\nu}_{\rm M}$ and not just to its $(0,0)$ component, it is not sufficient to only address the vacuum energy density problem, one has to deal with the vacuum pressure as well. There are thus two vacuum problems that need to be addressed, and not just one, with the vacuum zero-point fluctuation problem being quite distinct from the cosmological constant problem itself. Moreover, since it is the hallmark of gravity that gravity couple to energy density itself and not to energy density difference, one cannot ignore zero-point infinities or normal-order them away. Rather, one has to explicitly take them into consideration, and as we shall show below, we will ultimately be able to cancel them via an analogous zero-point structure that exists in the gravity sector itself.

For the purposes of parameterizing the divergences in (3) so that we can explicitly identify what it is that we will need to cancel, it is convenient to introduce a non-covariant momentum cut-off, with $\rho_{\rm M}$ and $p_{\rm M}$ then being given by

$$\begin{aligned}\rho_{\rm M} &= -\frac{\hbar}{4\pi^2}\left(K^4+\frac{m^2K^2}{\hbar^2}-\frac{m^4}{4\hbar^4}\ln\left(\frac{4\hbar^2K^2}{m^2}\right)+\frac{m^4}{8\hbar^4}\right),\\ p_{\rm M} &= -\frac{\hbar}{12\pi^2}\left(K^4-\frac{m^2K^2}{\hbar^2}+\frac{3m^4}{4\hbar^4}\ln\left(\frac{4\hbar^2K^2}{m^2}\right)-\frac{7m^4}{8\hbar^4}\right).\end{aligned} \tag{5}$$

We thus encounter a mass-independent quartic divergence and mass-dependent quadratic and logarithmic divergences. Hence, mass generation will not merely change the vacuum energy density and pressure, it will change them by infinite amounts, an effect that we will take care of below by having the mass generation be associated with an induced cosmological constant that will equally be divergent.

If one wishes to define the integral in (3) via the use of a set of covariant Pauli-Villars regulator masses M_i with parameters η_i, the choice $1+\sum\eta_i=0$, $m^2+\sum\eta_iM_i^2=0$, $m^4+\sum\eta_iM_i^4=0$ will not only then lead to finite regulated $\rho_{\rm REG}$ and $p_{\rm REG}$, it will give them the values

$$\rho_{\rm REG}=-p_{\rm REG}=-\frac{\hbar}{16\pi^2}\left(m^4\ln m^2+\sum\eta_iM_i^4\ln M_i^2\right). \tag{6}$$

The regulation procedure will thus make $\rho_{\rm REG}+p_{\rm REG}$ be equal to zero, just as required of a cosmological constant term, with a regulated $\langle\Omega|T^{\mu\nu}_{\rm M}|\Omega\rangle$ then behaving as $-\rho_{\rm REG}\eta^{\mu\nu}$. Thus we see that it is only the finite part of $\langle\Omega|T^{\mu\nu}_{\rm M}|\Omega\rangle$ that will behave like a cosmological constant term (though it would be a huge one if (6) is any indicator), while its infinite part could have a more complicated tensor structure even if the vacuum is Lorentz invariant. In general then we have to deal not just with a vacuum energy density problem, but with a vacuum pressure problem as well. Moreover, while different choices of energy-momentum tensor can be made that all lead to the same total energy (such as the canonical one or the Belinfante one for instance), these various choices lead to differing expressions for the pressure. Thus to correctly define the pressure, one must define the energy-momentum tensor as the functional variation with respect to the metric of a general-coordinate invariant action, just as in (2).

As we see from the structure of (5), in a massless theory there will only be a quartically divergent term. We can cancel this term by having some additional quartic divergence with the opposite sign. Moreover, this is precisely how supersymmetry does the cancellation because the vacuum energy densities of fermionic and bosonic fields have opposite signs. However, for this cancellation mechanism to be maintained following mass generation, the fermions and their bosonic superpartners would need to acquire degenerate masses (which experimentally we know not to be the case), since there would otherwise be uncanceled quadratic and logarithmic divergences. And even if these uncanceled terms are to be associated with the cut-off scale of some low energy effective theory, they would still make a huge contribution in (1). Thus using either supersymmetry or regulators as in (6), one can anticipate an eventual huge effective cosmological constant term, and it has yet to be shown that this is not in fact the case in theories based on (1).

One remaining option is to cancel the quartic divergences of the massless fermionic theory by conformal symmetry instead, and here the needed bosonic contribution would come from the gravitational sector as per (2) rather than (1). Moreover, as we show below, this particular cancellation mechanism will not be destroyed by mass generation. However, if we try to regulate the fermion vacuum energy as in (6), while such regulators would not violate Lorentz invariance, their masses would violate conformal invariance and lead to conformal anomalies. To avoid any such anomalies we must not introduce any regulators. Rather, we must have gravity itself do the cancellation just as in (2). Thus we need to cancel $\langle\Omega|T_{\rm M}^{\mu\nu}|\Omega\rangle$ mode by mode so as to eliminate the need to do any integration over modes. As we will see, this is precisely what quantizing gravity will guarantee us, with (2) fixing the normalization of the gravity sector commutators so as to ensure that the needed cancellations explicitly occur mode by mode. Moreover, in the absence of anomalies, in the following we will be able to maintain the tracelessness of both the matter and the gravitational energy-momentum tensors that is required by their underlying conformal structure.

3. TWO-DIMENSIONAL QUANTUM EINSTEIN GRAVITY

In analog to (5), for a free massless fermion in a two-dimensional flat spacetime, a canonical quantization of the form $\{\psi_\alpha(x,t),\psi_\beta^\dagger(x',t)\}=\delta(x-x')\delta_{\alpha,\beta}$ leads to a perfect fluid form for $\langle\Omega|T_{\rm M}^{\mu\nu}|\Omega\rangle$ with

$$\rho_{\rm M}=p_{\rm M}=-\frac{\hbar K^2}{2\pi}, \tag{7}$$

and a two-dimensional trace $p_{\rm M}-\rho_{\rm M}$ that vanishes. The task for two-dimensional gravity is thus to cancel this quadratic divergence. Now while the two-dimensional Einstein-Hilbert action $I_{\rm GRAV}=-(1/2\kappa_2^2)\int d^2x(-g)^{1/2}R^\alpha{}_\alpha$ is conformal invariant (Newton's constant κ_2^2 being dimensionless in two dimensions), this action has the property that as a classical action, it is a total divergence in any gravitational path (the two-dimensional Gauss-Bonnet theorem). Consequently the classical Einstein tensor will vanish identically for any choice of classical metric $g^{\mu\nu}(x)$ whatsoever. Moreover, with the quantum

theory being given as a path integral over classical gravitational metrics, the path integral is trivial, and quantum-mechanically there is no gravitational scattering. Thus the classical theory does not exist, and quantum radiative corrections do not exist either.

Despite these concerns, they do not mean that the quantum theory is completely empty, since while the path integral gives correlation functions, it does not give the vacuum energy. Rather, it only gives the normal-ordered finite piece of the vacuum energy, since the path integral would otherwise not exist. Thus in two-dimensional quantum Einstein gravity there could still be, and there in fact is, a non-vanishing vacuum energy. To see how it can arise, we note that in generic relations such as $A\partial_\mu B + B\partial_\mu A = \partial_\mu(AB) + [B, \partial_\mu A]$, because of ordering a function that would be a total divergence classically need not be one quantum-mechanically. Hence in the presence of quantum ordering the theory need no longer be completely trivial. To take ordering into account, we need to specify a choice of ordering, and in order to enforce symmetry of the Ricci tensor, we once and for all define geometric tensors according to the ordering sequencing $R_{\lambda\mu\nu\kappa} = (1/2)[\partial_\kappa\partial_\mu g_{\lambda\nu} - \partial_\kappa\partial_\lambda g_{\mu\nu} - \partial_\nu\partial_\mu g_{\lambda\kappa} + \partial_\nu\partial_\lambda g_{\mu\kappa}) + g_{\eta\sigma}(\Gamma^\eta_{\nu\lambda}\Gamma^\sigma_{\mu\kappa} - \Gamma^\eta_{\kappa\lambda}\Gamma^\sigma_{\mu\nu}]$, $\Gamma^\alpha_{\mu\kappa} = (1/2)g^{\alpha\beta}(\partial_\mu g_{\beta\kappa} + \partial_\kappa g_{\beta\mu} - \partial_\beta g_{\mu\kappa})$, $R_{\mu\kappa} = (1/2)[g^{\nu\lambda}R_{\lambda\mu\nu\kappa} + g^{\nu\lambda}R_{\lambda\kappa\nu\mu}]$, $R^\alpha{}_\alpha = g^{\mu\kappa}R_{\mu\kappa}$, and $G_{\mu\kappa} = R_{\mu\kappa} - (1/2)g_{\mu\kappa}R^\alpha{}_\alpha$.

If we perturb to second order around flat spacetime according to $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$, $g^{\mu\nu} = \eta^{\mu\nu} - h^{\mu\nu} + h^\mu{}_\sigma h^{\sigma\nu}$, then we find that the first order $G_{\mu\nu}(1)$ vanishes identically (as it of course must since there is no ordering issue in first order and $G_{\mu\nu}$ already vanishes classically). However, for our choice of ordering, in second order we obtain

$$\begin{aligned} G_{00}(2) &= \frac{1}{8}[\partial_0 h_{00}, 2\partial_1 h_{01} - \partial_0 h_{11}] + \frac{1}{8}[\partial_1 h_{11}, 2\partial_0 h_{01} - \partial_1 h_{00}] = G_{11}(2), \\ G_{01}(2) &= \frac{1}{8}[\partial_1 h_{00}, 2\partial_1 h_{01} - \partial_0 h_{00}] + \frac{1}{8}[\partial_0 h_{11}, 2\partial_0 h_{01} - \partial_1 h_{11}], \end{aligned} \tag{8}$$

with the two-dimensional $G^{\mu\nu}(2)$ automatically being traceless, just as it should be in a conformal theory. As we thus see, the quantum $G_{\mu\nu}(2)$ is given by a set of commutator terms, terms that would vanish classically but not quantum-mechanically, to thus make the quantum theory non-trivial despite the triviality of the classical theory. Given (8) we can evaluate the various components of the covariant derivative of $G^{\mu\nu}(2)$, to obtain

$$\begin{aligned} \partial_\mu G^\mu{}_0(2) &= \frac{1}{4}[\nabla^2 h_{00}, \partial_1 h_{01}] - \frac{1}{4}[\nabla^2 h_{01}, \partial_1 h_{11}] - \frac{1}{8}[\partial_0\partial_1(h_{00} + h_{11}), \partial_1 h] \\ &\quad - \frac{1}{4}[\partial_0\partial_1 h_{01}, \partial_0 h] + \frac{1}{4}[\partial_1^2 h_{01}, \partial_1 h] + \frac{1}{8}[\nabla^2 h, \partial_0 h_{00}] + \frac{1}{8}[\partial_0^2 h_{00} + \partial_1^2 h_{11}, \partial_0 h], \\ \partial_\mu G^\mu{}_1(2) &= \frac{1}{4}[\nabla^2 h_{11}, \partial_0 h_{01}] - \frac{1}{4}[\nabla^2 h_{01}, \partial_0 h_{00}] - \frac{1}{8}[\partial_0\partial_1(h_{00} + h_{11}), \partial_0 h] \\ &\quad - \frac{1}{4}[\partial_0\partial_1 h_{01}, \partial_1 h] + \frac{1}{4}[\partial_0^2 h_{01}, \partial_0 h] - \frac{1}{8}[\nabla^2 h, \partial_1 h_{11}] + \frac{1}{8}[\partial_0^2 h_{00} + \partial_1^2 h_{11}, \partial_1 h], \end{aligned} \tag{9}$$

where $\nabla^2 = -\partial_0^2 + \partial_1^2$, $h = \eta^{\mu\nu}h_{\mu\nu} = -h_{00} + h_{11}$. Since $G_{\mu\nu}(1)$ vanishes trivially, there is no first-order equation of motion that would force $\partial_\mu G^{\mu\nu}(2)$ to vanish identically. Consequently, the Bianchi identity of classical gravity does not automatically have to hold quantum-mechanically. Nonetheless, because of the matter field wave equation, the

matter field energy-momentum tensor is covariantly conserved. Thus from the quantum-mechanical field equation

$$\frac{1}{\kappa_2^2}G^{\mu\nu}+T_{\rm M}^{\mu\nu}=0, \tag{10}$$

it follows that $\partial_\mu G^{\mu\nu}(2)$ does vanish after all. As we thus see, unlike the standard classical situation, in the quantum theory $G^{\mu\nu}(2)$ is only covariantly conserved on the stationary path and not on the arbitrary one.

Since $T_{\rm M}^{\mu\nu}$ is of order $\hbar$ in (7), to satisfy (10) to this order we must take $h_{\mu\nu}$ to be of order $h^{1/2}$, with (10) then fixing the gravitational commutators in $G^{\mu\nu}(2)$ to be of order $\hbar$. It is thus the quantization of the matter field that forces the quantization of the gravitational field with $G^{\mu\nu}$ not being able to vanish once $T_{\rm M}^{\mu\nu}$ does not. Through order $\hbar$ we can take $T_{\rm M}^{\mu\nu}$ to have the value it would have in flat space, with curvature corrections to it only appearing in higher order in $\hbar$. However, the quantum $G^{\mu\nu}$ is non-trivial already in order $\hbar$. Finally, with the vanishing of $\partial_\mu G^{\mu\nu}(2)$ as now enforced by the vanishing of $\partial_\mu T_{\rm M}^{\mu\nu}(2)$, from (10) we see that the components of $h_{\mu\nu}$ can be taken to obey

$$\nabla^2 h_{00}=0, \qquad \nabla^2 h_{01}=0, \qquad \nabla^2 h_{11}=0, \qquad h=-h_{00}+h_{11}=0, \tag{11}$$

to thus obey a massless wave equation after all, with the trace of $h_{\mu\nu}$ vanishing just as is to be expected in a conformal theory.

Given (11), we can expand the quantum fields in a complete basis of plane waves with $k^\mu=(\omega_k,k)$ where $\omega_k=|k|$. We thus introduce creation and annihilation operators and set

$$\begin{aligned} h_{00}(x,t) &= \kappa_2\hbar^{1/2}\int\frac{dk}{(2\pi)^{1/2}(2\omega_k)^{1/2}}\left[A(k)e^{i(kx-\omega_k t)}+C(k)e^{-i(kx-\omega_k t)}\right]=h_{11}(x,t),\\ h_{01}(x,t) &= \kappa_2\hbar^{1/2}\int\frac{dk}{(2\pi)^{1/2}(2\omega_k)^{1/2}}\left[B(k)e^{i(kx-\omega_k t)}+D(k)e^{-i(kx-\omega_k t)}\right]. \end{aligned} \tag{12}$$

We now introduce a vacuum for the Hilbert space, and as usual take the two positive frequency operators $A(k)$ and $B(k)$ to annihilate the right vacuum $|\Omega\rangle$, and the two negative frequency operators $C(k)$ and $D(k)$ to annihilate the left vacuum $\langle\Omega|$. In addition we require that the vacuum expectation values of the commutators $[C(k),B(k')]$ and $[A(k),D(k')]$ be given as

$$\begin{aligned} \langle\Omega|[C(k),B(k')]|\Omega\rangle &= -\langle\Omega|B(k)C(k)|\Omega\rangle\delta(k-k')=-f_{BC}(k)\delta(k-k'),\\ \langle\Omega|[A(k),D(k')]|\Omega\rangle &= \langle\Omega|A(k)D(k)|\Omega\rangle\delta(k-k')=f_{AD}(k)\delta(k-k'). \end{aligned} \tag{13}$$

Finally, with (10) taking the form $\langle\Omega|G^{00}(2)|\Omega\rangle/\kappa_2^2=(\hbar/8\pi)\int_{-\infty}^{\infty}dkk[f_{BC}(k)-f_{AD}(k)]=-\rho_{\rm M}$, $\langle\Omega|G^{01}(2)|\Omega\rangle/\kappa_2^2=(\hbar/8\pi)\int_{-\infty}^{\infty}dk\omega_k[f_{BC}(k)-f_{AD}(k)]=0$, we find that the c-number functions $f_{BC}(k)$ and $f_{AD}(k)$ have to obey just the one relation

$$k[f_{BC}(k)-f_{AD}(k)]=4\omega_k=4|k|. \tag{14}$$

With this one quantization condition, we achieve our primary purpose of showing how the quadratically divergent zero-point fluctuations of the gravitational and matter fields

mutually cancel each other identically, just as desired. Moreover, through its coupling to the quantized fermionic field the gravitational field commutators are forced to obey (14), with gravity not needing any independent quantization of its own.

4. MASS GENERATION AND THE COSMOLOGICAL CONSTANT

For our purposes here we can take the fermionic action to be of the flat spacetime form $I_{\rm M} = -\int d^2x[i\hbar\bar{\psi}\gamma^\mu\partial_\mu\psi - (g/2)(\bar{\psi}\psi)^2]$. With the four-Fermi coupling constant g being dimensionless in two dimensions, $I_{\rm M}$ is conformal invariant. Consequently, as well as being covariantly conserved, its energy-momentum tensor $T^{\mu\nu}_{\rm M} = i\hbar\bar{\psi}\gamma^\mu\partial^\nu\psi - \eta^{\mu\nu}(g/2)[\bar{\psi}\psi]^2$ is traceless in solutions to the equation of motion. In the Nambu-Jona-Lasinio mean-field, Hartree-Fock approximation one looks for self-consistent, translation invariant states $|S\rangle$ in which $\langle S|\bar{\psi}\psi|S\rangle = \langle S|\psi^\dagger\gamma^0\psi|S\rangle = im/g$ and $\langle S|(\bar{\psi}\psi - im/g)^2|S\rangle = 0$. (With our choice of $\eta^{00} = -1$, γ^0 is pure imaginary.) In such states the fermion equation of motion takes the form $i\hbar\gamma^\mu\partial_\mu\psi - im\psi = 0$ and the mean-field fermion energy-momentum tensor $T^{\mu\nu}_{\rm MF}$ takes the form

$$\langle S|T^{\mu\nu}_{\rm MF}|S\rangle = \langle S|i\hbar\bar{\psi}\gamma^\mu\partial^\nu\psi|S\rangle + \frac{m^2}{2g}\eta^{\mu\nu}, \tag{15}$$

with the mean-field approximation preserving tracelessness. In conformal invariant theories then, we see that, just as noted in [1], one can have mass generation without the trace needing to be non-zero. With the emergence of the $(m^2/2g)\eta^{\mu\nu}$ term in (15), we see that dynamical mass generation induces a mean-field cosmological constant term $\Lambda_{\rm MF} = -m^2/2g$, and that with this $\Lambda_{\rm MF}$ we can write $\langle S|T^{\mu\nu}_{\rm MF}|S\rangle$ as

$$\langle S|T^{\mu\nu}_{\rm MF}|S\rangle = (\rho_{\rm MF} + p_{\rm MF})U^\mu U^\nu + p_{\rm MF}\eta^{\mu\nu} - \Lambda_{\rm MF}\eta^{\mu\nu}, \tag{16}$$

where

$$\begin{aligned}
\rho_{\rm MF} &= -\frac{\hbar}{2\pi}\left[K^2 + \frac{m^2}{2\hbar^2} + \frac{m^2}{2\hbar^2}\ln\left(\frac{4\hbar^2K^2}{m^2}\right)\right], \\
p_{\rm MF} &= -\frac{\hbar}{2\pi}\left[K^2 + \frac{m^2}{2\hbar^2} - \frac{m^2}{2\hbar^2}\ln\left(\frac{4\hbar^2K^2}{m^2}\right)\right], \quad \Lambda_{\rm MF} = \frac{m^2}{4\pi\hbar}\ln\left(\frac{4\hbar^2K^2}{m^2}\right),
\end{aligned} \tag{17}$$

and where the expression for $\Lambda_{\rm MF} = -m^2/2g$ is recognized as the gap equation $m = 2\hbar K e^{\pi\hbar/g}$.

In (17) we see that the mass-independent quadratic divergences in $\rho_{\rm MF}$ and $p_{\rm MF}$ have been augmented by mass-dependent logarithmic ones, while the induced $\Lambda_{\rm MF}$ is logarithmically divergent (i.e. not finite). However, since $\langle S|T^{\mu\nu}_{\rm MF}|S\rangle$ is traceless, the various terms in (17) obey $p_{\rm MF} - \rho_{\rm MF} - 2\Lambda_{\rm MF} = 0$, with all the various divergences canceling each other in the trace, just as noted in [1, 6]. Given this cancellation, we can now use the trace condition to eliminate $\Lambda_{\rm MF}$ and rewrite (16) as

$$\langle S|T^{\mu\nu}_{\rm MF}|S\rangle = (\rho_{\rm MF} + p_{\rm MF})\left[U^\mu U^\nu + \frac{1}{2}\eta^{\mu\nu}\right], \quad \rho_{\rm MF} + p_{\rm MF} = -\frac{\hbar}{\pi}\left(K^2 + \frac{m^2}{2\hbar^2}\right), \tag{18}$$

with the logarithmic divergences associated with the mass-induced readjustment of ρ_{MF} and p_{MF} having disappeared completely. Finally, to cancel the remaining quadratic divergence and finite part in (18), we proceed just as in the massless fermion case, only with (14) having to be replaced by

$$k[f_{BC}(k) - f_{AD}(k)] = 4\left[(k^2 + m^2/\hbar^2)^{1/2} - \frac{m^2}{2\hbar^2(k^2 + m^2/\hbar^2)^{1/2}}\right]. \qquad (19)$$

In (19) we note that even though the gravitational field is massless and still obeys (11), its quantization condition depends on the mass of the fermion, a reflection of the fact that it is only through the quantization of the fermionic field that the gravitational field is quantized in the first place.

As such, the above analysis shows how the vacuum contribution to the cosmological constant is completely taken care of by the zero-point contributions. However, there is one further concern that still needs to be addressed, as there is a further contribution to the cosmological constant term matrix element, namely that associated with occupying not just the vacuum $|S\rangle$ but the one-particle excitations as well. In such states $|C\rangle$ the quantity $im(x)/g = \langle C|\bar{\psi}(x)\psi(x)|C\rangle$ can typically acquire a spacetime dependence. However, as noted in [4], the spatial dependence will asymptote to the constant vacuum value while the time dependence will redshift. The cosmological term needed for cosmology is thus given not by the vacuum contribution itself but by the spatial departure from it, i.e. by $m^2(x)/2g - m^2/2g$, as redshifted to the current era. Such an effective cosmological term would not at all be constrained to be as large as the vacuum value, but its actual value still needs to be determined.

5. ACKNOWLEDGMENTS

This paper is based in part on a presentation made by the author at the International Conference on Two Cosmological Models, Universidad Iberoamericana, Mexico City, November 2010. The author wishes to thank Dr. J. Auping and Dr. A. V. Sandoval for the kind hospitality of the conference. The author also wishes to thank Dr. D. J. Gross for helpful comments.

REFERENCES

1. P. D. Mannheim, *Prog. Part. Nucl. Phys.* **56**, 340 (2006).
2. C. M. Bender and P. D. Mannheim, *Phys. Rev. Lett.* **100**, 110402 (2008).
3. C. M. Bender and P. D. Mannheim, *Phys. Rev. D* **78**, 025022 (2008).
4. P. D. Mannheim, *Gen. Relativ. Gravit.* **43**, 703 (2011).
5. P. D. Mannheim, *Making the case for conformal gravity*, arXiv:1101.2186v1 [hep-th].
6. P. D. Mannheim, *Dynamical symmetry breaking and the cosmological constant problem*, Proceedings of the 34th International Conference in High Energy Physics (ICHEP08), Philadelphia, 2008, eConf C080730. (arXiv:0809.1200 [hep-th])

Towards physical cosmology:

Geometrical interpretation of Dark Energy, Dark Matter and Inflation without Fundamental Sources

Thomas Buchert

Université Lyon 1, Centre de Recherche Astrophysique de Lyon, CNRS UMR 5574
9 Avenue Charles André, F–69230 Saint–Genis–Laval, France

Abstract. We outline the key–steps towards the construction of a physical, fully relativistic cosmology, in which we aim to trace Dark Energy and Dark Matter back to physical properties of space. The influence of inhomogeneities on the effective evolution history of the Universe is encoded in backreaction terms and expressed through spatially averaged geometrical invariants. These are absent and interpreted as missing dark fundamental sources in the standard model. In the inhomogeneous case they can be interpreted as energies of an emerging scalar field (the morphon). These averaged invariants vanish for a homogeneous geometry, where the morphon is in an unstable equilibrium state. If this state is perturbed, the morphon can act as a classical inflaton in the Early Universe and its de–balanced energies can mimic the dark sources in the Late Universe, depending on spatial scale as Dark Energy or as Dark Matter, respectively. We lay down a line of arguments that is qualitatively conclusive, and we outline open problems of quantitative nature, related to the interpretation of observations.

Keywords: Inhomogeneous Cosmology, Dark Energy, Dark Matter, Inflation
PACS: 98.80.-k, 98.80.Cq, 95.35.+d, 95.36.+x, 98.80.Es, 98.80.Jk,04.20.-q

GENERAL RELATIVITY AND COSMOLOGY

The foliation issue and the notion of an effective cosmology

The homogeneous–isotropic standard model of cosmology, being itself a particular solution of Einstein's general theory of relativity, does by far not exploit the degrees of freedom inherent in the geometry as a dynamical variable. It is this richer tone of general relativity – as compared to the Newtonian theory – that opens the possibility to generalize cosmological models, notably by including inhomogeneous structure also in the geometrical variables. There are several guidelines to be emphasized in such a generalization: firstly, a cosmology is thought of as an evolving space section that implies the need to speak of a foliated space time, introducing four degrees of freedom (the lapse and shift functions in an ADM setting). This necessarily implies, on general grounds, a breaking of four–dimensional covariance. This fact should not be confused with coordinate– or gauge–dependence of the resulting cosmological equations and variables, however. Secondly, a cosmology purports an effective point of view in the sense that the evolving spatially inhomogeneous variables are thought of as being "averaged

over" in a way that has to be specified. We aim at a description that only implicitly refers to a metric. However, if a metric is to be specified, a cosmological metric is then to be considered as an effective, "smoothed out" or *template metric*, being not necessarily a solution of the equations of general relativity. Finally, a *physical* cosmology should be characterized by such an effective evolution model, an effective metric to provide the distance scale for the interpretation of observations, or alternatively an evolution model for average characteristics on the light cone, together with a set of initial data. These latter are to be related to physical properties of fundamental sources, but also to the geometrical data at some initial time (effective, i.e "averaged" quantities of known energy sources, intrinsic and extrinsic curvature). This latter clearly emphasizes the absence of any phenomenological parameters. Those would just parametrize our physical ignorance. All these points will be made explicit in what follows.

The dark side of the standard model: postulated sources and proposed solutions

The high level of idealization of the geometrical properties of space in the standard model leads to the need of postulating sources that would generate "on average" a strictly, i.e. globally and locally, homogeneous geometry. It is here where a considerable price has to be paid for a model geometry that obviously is not enough to meet physical reality, unless we really believe that we can find the missing sources: 96 percent of the energy content is missing in the form of a) a postulated source acting attractive like matter, so–called Dark Matter ($\cong$ 23 percent) and b) a postulated source acting repulsive, so–called Dark Energy ($\cong$ 73 percent). Evidence for the former does indeed come from various scales (galaxy halos, clusters and cosmological, see e.g. [63]), while evidence for the latter only comes from the apparent magnitude of distant supernovae (see [42, 35, 30] for the latest data) that, if interpreted within standard model distances, would need an accelerating model. In the simplest case this volume acceration is achieved by a homogeneous–isotropic cosmology with a cosmological constant. It should be emphasized that when we speak of evidence, we already approach this evidence with model priors [36, 68, 70]. Keeping this idealization for the geometry of the cosmological model for example, one has to conjecture fundamental fields in proportion to the missing dark components on cosmological scales. The search for these fields is one major research direction in modern cosmology.

Another huge effort is directed towards a generalization of the underlying theory of gravitation. While this would generalize the geometry of the model, it is not clear why all these efforts go into a generalization of general relativity and not into the generalization of the cosmological model within general relativity. There are certainly good lines of arguments and various motivations in particle physics and quantum gravity to go beyond the theory of Einstein (for reviews see [25], [66]), but the "dark problem" may be first a classical one.

Looking at generalizations of the standard model *within* general relativity can be identified as a third research direction to which we dedicate our attention here. In light of current efforts it is to be considered conservative, since it does not postulate new

fundamental fields and it does not abandon a well–tested theory of gravitation [13], [56], [39] (for reviews on the physical basis of this third approach see [8, 18] and [57]). Among the works in this latter field, research that analyzes spherically symmetric exact solutions has been meanwhile developed to some depth, and has determined the constraints, necessary to explain Dark Energy, on a postulated observer's position within a large–scale void (see [29, 4, 22, 3] and references therein).

Fictitious and physical backgrounds

Perhaps a reason for not questioning the standard model geometry within general relativity and to go for the search for fundamental fields or for generalizations of the laws of gravitation is the following belief: effectively, i.e. "on average", the model geometry has to be *homogeneous*, since structures should be "averaged over". Then, due to observational facts on large scales (the high degree of isotropy of the Cosmic Microwave Background, if the dipole is completely eliminated due to our proper motion with respect to an idealized exactly isotropic light sphere), and first principle priors (the *strong cosmological principle* that requires the universe model to look the same in all directions), the model geometry is taken to be *locally isotropic*.

Taking this reasoning at face value we must note two points: the notions of homogeneity and isotropy in the standard model are both too strong to be realistic: firstly, local isotropy implies a model that is locally and globally homogeneous, i.e. despite the conjecture that the homogeneous model describes the inhomogeneous Universe "on average", this *strict homogeneity* does not account for the fact that any averaging procedure, in one way or another, would introduce a *scale–dependence* of the averaged (homogeneous) variables [28]. This scale–dependence, inherent in any physical averages, is suppressed. Even if a large *scale of homogeneity* exists (we may call this *weak homogeneity principle*), the model is in general scale–dependent on scales below this homogeneity scale [69]. The same is true for isotropy: while the averaged model may be highly isotropic on large scales, a realistic distribution on smaller scales is certainly not (we may call this *weak isotropy principle*). Correspondingly, a *weak cosmological principle* would be enough to cover the reality needs while still facing observational evidence on large scales.

We may summarize the above thoughts by noting that, on large scales, a homogeneous–(almost)isotropic *state* does not necessarily correspond to a homogeneous–(almost)isotropic *solution* of Einstein's equations. These former states are the averages over fluctuating fields and it is only to be expected that the state coincides with a strictly homogeneous solution in the case of absence of fluctuations. In other words, looking at fluctuations first requires to establish the average distribution. Only then the notion of a *background* makes physical sense [41]. Current cosmological structure formation models, perturbation theories or N–body simulations, are constructed such that the average vanishes on the background of a homogeneous–isotropic *solution* [7]. A such chosen reference background may be a *fictitious background*, since it arises by construction rather than derivation. On the contrary, a *physical background* is one that corresponds to the average (whose technical implementation has to be

specified, and which is nontrivial if tensorial quantities like the geometry have to be "averaged"). A sound implementation of a physical background will be a statistical background where not only solutions but ensembles of solutions are averaged. Having specified such an averaging procedure, a physical cosmological model may then be defined as an evolution model for the average distribution. Despite these remarks it is of course possible that the homogeneous solution forms at the same time the average. A well–known example is Newtonian cosmology [7]. It is also conceivable that the homogeneous solution provides, in some spatial and temporal regimes, a good approximation for the average. Still, it is important to consider perturbations on the correct background solution [40].

SCALAR FIELD MODELS AND THE MORPHON

Effective evolution of inhomogeneous universe models

Taking the point of view of generalizing the cosmological model within general relativity by abandoning the strong cosmological principle (strict homogeneity and isotropy on all scales) and replacing it by the weak cosmological principle (existence of a homogeneity scale and restriction to effective states that are almost isotropic on the scale of homogeneity) leads us to a "rewriting of the rules" to build the cosmological model. We shall consider the rules that led to the standard model of cosmology and replace them by their more general counterparts. It follows a basically similar framework that displays, however, a signature of inhomogeneity through the occurence of so–called backreaction terms and through a manifest scale–dependency. We shall not introduce new principles or assumptions, apart from the above outlined relaxation of the cosmological principle. We shall restrict ourselves to the simplest case of an irrotational dust model (for generalizations of the dust model [9] with non–constant lapse function see [10], and for additionally non–vanishing shift see [6, 5, 44, 33]).

• As in the standard model we introduce a foliation of space time into flow–orthogonal hypersurfaces. We generalize the notion of *Fundamental Observers* to those that are in free fall also in the general space time. Although, as in the standard model, this setup depends on the chosen foliation, we presume that this choice is unique as it prefers the fundamental observers against observers that may be accelerated with respect to the hypersurfaces. A general inhomogeneous hypersurface – contrary to the homogeneous case – will, in this setting, unavoidably run into singularities in the course of evolution. This is to be expected in a given range of spatial and temporal scales, since we are treating the matter model as *dust*. This is not a problem of the chosen foliation, but a problem of the matter model that has to be generalized, if small–scale structure formation has to remain regular, and this can be achieved by the inclusion of velocity dispersion and vorticity.

• As in the standard model we confine ourselves to scalar quantities. We replace, however, the homogeneous quantities by their spatial averages, e.g. the homogeneous density $\rho_H(t)$ is replaced by $\langle\rho\rangle_{\mathscr{D}}(t)$ for the inhomogeneous density ρ that is volume–averaged over some compact domain $\mathscr{D}$. We realize the averaging operation by a mass–preserving

Riemannian volume average. In some mathematical disciplines and in statistical averages at one instant of time, it may be more convenient to introduce a volume–preserving averager, but thinking of an averaging domain that is as large as the homogeneity scale we have to preserve mass rather than volume. Furthermore, the average is performed with respect to the above–defined *Fundamental Observers*. Spatially averaging a scalar $\Psi(t,X^i)$, as a function of Gaussian coordinates X^i and a synchronizing time t, is defined as:

$$\langle\Psi(t,X^i)\rangle_{\mathscr{D}} := \frac{1}{V_{\mathscr{D}}}\int_{\mathscr{D}} \Psi(t,X^i)\, d\mu_g \quad , \tag{1}$$

with the Riemannian volume element $d\mu_g := \sqrt{g}d^3X$, $g := \det(g_{ij})$, and the volume of an arbitrary compact domain, $V_{\mathscr{D}}(t) := \int_{\mathscr{D}}\sqrt{g}d^3X$. Note that within a more general setup that includes lapse and shift functions, we would have to consider the question whether the locally lapsed time is replaced by a global "averaged time" that would involve an average over the lapse function. Here, the dust cosmology is already synchronous, so that this question does not arise. Note furthermore, that the building of averages is done in the inhomogeneous geometry. The averages functionally depend on the inhomogeneous metric, but this latter needs not to be specified. We may talk of a kinematical averaging that does not deform the geometry, i.e. that does not change the physical properties of the inhomogeneous space time. For other strategies, see [28], and references therein, as well as Section IV.

• We generalize the kinematical laws of the standard model a) for the volume expansion (the Hamiltonian constraint in the ADM formulation of general relativity) and b) for the volume acceleration (Raychaudhuri's equation in the ADM formulation of general relativity) by dropping the symmetry assumption of local isotropy. The general equations are then volume–averaged, leading to the following general volume expansion and volume acceleration laws (for a volume scale factor, defined by $a_{\mathscr{D}}(t) := (V_{\mathscr{D}}(t)/V_{\mathscr{D}}(t_i))^{1/3}$; the overdot denotes partial time–derivative, which is the covariant time–derivative here) [9]:

$$3\frac{\ddot{a}_{\mathscr{D}}}{a_{\mathscr{D}}} = -4\pi G\langle\rho\rangle_{\mathscr{D}} + Q_{\mathscr{D}} + \Lambda \quad ; \quad 3H^2_{\mathscr{D}} + \frac{3k_{\mathscr{D}}}{a^2_{\mathscr{D}}} = 8\pi G\langle\rho\rangle_{\mathscr{D}} - \frac{1}{2}W_{\mathscr{D}} - \frac{1}{2}Q_{\mathscr{D}} + \Lambda\,, \tag{2}$$

where $H_{\mathscr{D}}$ denotes the domain dependent Hubble rate $H_{\mathscr{D}} = \dot{a}_{\mathscr{D}}/a_{\mathscr{D}} = -1/3\langle K\rangle_{\mathscr{D}}$, K is the trace of the extrinsic curvature of the embedding of the hypersurfaces into the space time, K_{ij}, and Λ the cosmological constant. The *kinematical backreaction* $Q_{\mathscr{D}}$ is composed of averaged extrinsic curvature invariants, while $W_{\mathscr{D}}$ is an averaged intrinsic curvature invariant that describes the deviation of the average of the full (three–dimensional) Ricci scalar curvature R from a constant–curvature model,

$$Q_{\mathscr{D}} := \left\langle K^2 - K^i_{\ j}K^j_{\ i}\right\rangle_{\mathscr{D}} - \frac{2}{3}\langle K\rangle^2_{\mathscr{D}} \quad ; \quad W_{\mathscr{D}} := \langle R\rangle_{\mathscr{D}} - \frac{6k_{\mathscr{D}}}{a^2_{\mathscr{D}}}\,. \tag{3}$$

The kinematical backreaction $Q_{\mathscr{D}}$ can also be expressed in terms of kinematical invariants, where the extrinsic curvature is interpreted actively in terms of (minus) the expansion tensor:

$$Q_{\mathscr{D}} := \frac{2}{3}\left(\left\langle\theta^2\right\rangle_{\mathscr{D}} - \langle\theta\rangle^2_{\mathscr{D}}\right) - 2\left\langle\sigma^2\right\rangle_{\mathscr{D}}\,, \tag{4}$$

where θ is the local expansion rate and $\sigma^2 := 1/2\sigma_{ij}\sigma^{ij}$ is the squared rate of shear. Note that $H_{\mathscr{D}}$ is now defined as $H_{\mathscr{D}} = 1/3\,\langle\theta\rangle_{\mathscr{D}}$. $Q_{\mathscr{D}}$ appears as a competition term between the averaged variance of the local expansion rates, $\left\langle\theta^2\right\rangle_{\mathscr{D}} - \langle\theta\rangle^2_{\mathscr{D}}$, and the averaged square of the shear scalar $\left\langle\sigma^2\right\rangle_{\mathscr{D}}$ on the domain under consideration.

For a homogeneous domain the above backreaction terms $Q_{\mathscr{D}}$ and $W_{\mathscr{D}}$, being covariantly defined and gauge invariants in a perturbation theory on a homogeneous background solution, are zero. They encode the departure from homogeneity in a coordinate–independent way [46, 33].

The integrability conditions connecting the two Eqs. (2), assuring that the expansion law is the integral of the acceleration law, read:

$$\langle\rho\rangle^{\cdot} + 3H_{\mathscr{D}}\,\langle\rho\rangle_{\mathscr{D}} = 0 \quad ; \quad a_{\mathscr{D}}^{-2}(a_{\mathscr{D}}^2 W_{\mathscr{D}})^{\cdot} + a_{\mathscr{D}}^{-6}(a_{\mathscr{D}}^6 Q_{\mathscr{D}})^{\cdot} = 0\,. \tag{5}$$

While the mass conservation law for the dust is sufficient in the homogeneous case, there is a further equation connecting averaged intrinsic and extrinsic curvature invariants in the inhomogeneous case. The expressions in brackets are conformal invariants (for further details see [18].

The interpretation of these average equations as *generalized or evolving backgrounds* [18], [41] implies that the second conservation law describes an interaction between structure formation and background curvature. In the standard model this latter is absent and structures evolve independently of the background. This constant–curvature background furnishes the only solution of (5), in which structure formation decouples from the background (the expressions in brackets in the second conservation law are separately constant). Backreaction on such a fixed background decays in proportion to the square of the density and is unimportant in the Late Universe [9, 14, 18]. This degenerate case of a decoupled evolution explains the fact that in Newtonian and quasi–Newtonian models backreaction has no or little relevance [18]; in the Newtonian case [7], as well as in Newtonian [11, 38] and spatially flat, relativistic spherically symmetric dust solutions [52], $Q_{\mathscr{D}}$ vanishes. In models with homogeneous geometry and with periodic boundary conditions imposed on the inhomogeneities on some scale, the backreaction term is globally zero and describes cosmic variance of the kinematical properties.

Note here that, in general, a physical background "talks" with the fluctuations, and it is this coupling that gives rise to an instability of the constant–curvature backgrounds as we discuss below. The essential effect of backreaction models is not a large magnitude of $Q_{\mathscr{D}}$, but a dynamical coupling of a nonvanishing $Q_{\mathscr{D}}$ to the averaged scalar curvature deviation $W_{\mathscr{D}}$. This implies that the temporal behavior of the averaged curvature deviates from the behavior of a constant–curvature model. In concrete studies, as discussed further below, this turns out to be the major effect of backreaction, since it does not only change the kinematical properties of the cosmological model, but also the interpretation of observational data as we explain in Section IV.

Scalar field emerging from geometrical inhomogeneities

We rewrite the above set of spatially averaged equations together with their integrability conditions by appealing to the kinematical equations of the standard model, which will now be sourced by an effective perfect fluid energy–momentum tensor [10]:

$$3\frac{\ddot{a}_{\mathscr{D}}}{a_{\mathscr{D}}} = -4\pi G(\rho^{\mathscr{D}}_{\rm eff}+3p^{\mathscr{D}}_{\rm eff}) \,+\, \Lambda \quad ; \tag{6}$$

$$3H^2_{\mathscr{D}} - \frac{3k_{\mathscr{D}}}{a^2_{\mathscr{D}}} = 8\pi G\rho^{\mathscr{D}}_{\rm eff} \,+\, \Lambda \quad ; \tag{7}$$

$$\dot{\rho}^{\mathscr{D}}_{\rm eff} + 3H_{\mathscr{D}}\left(\rho^{\mathscr{D}}_{\rm eff}+p^{\mathscr{D}}_{\rm eff}\right) \,= 0\,, \tag{8}$$

where the effective densities are defined as

$$\begin{aligned} \rho^{\mathscr{D}}_{\rm eff} &:= \langle\rho\rangle_{\mathscr{D}} + \rho_\Phi \quad ; \quad \rho_\Phi \;:=\; -\frac{1}{16\pi G}Q_{\mathscr{D}} - \frac{1}{16\pi G}W_{\mathscr{D}}\,; \\ p^{\mathscr{D}}_{\rm eff} &:= p_\Phi \quad ; \quad p_\Phi \;:=\; -\frac{1}{16\pi G}Q_{\mathscr{D}} + \frac{1}{48\pi G}W_{\mathscr{D}}\,. \end{aligned} \tag{9}$$

In this form the effective equations suggest themselves to interpret the extra fluctuating sources in terms of a scalar field [15, 16], which refer to the inhomogeneities in geometrical variables. Thus, we choose to consider the averaged model as a (scale–dependent) "standard model" with matter source evolving in a *mean field* of backreaction terms. This scalar field we call the *morphon field*, since it captures the morphological (integral–geometrical [18]) signature of structure. (Note that in more general cases, involving lapse and shift functions, the structure of the scalar field theory suggested by the equations may no longer be a minimally coupled one.) We rewrite [16]:

$$\rho^{\mathscr{D}}_\Phi = \varepsilon\frac{1}{2}\dot{\Phi}^2_{\mathscr{D}} + U_{\mathscr{D}} \quad ; \quad p^{\mathscr{D}}_\Phi = \varepsilon\frac{1}{2}\dot{\Phi}^2_{\mathscr{D}} - U_{\mathscr{D}} \;, \tag{10}$$

where $\varepsilon = +1$ for a standard scalar field (with positive kinetic energy), and $\varepsilon = -1$ for a phantom scalar field (with negative kinetic energy; if ε is negative, a "ghost" can formally arise on the level of an effective scalar field, although the underlying theory does not contain one; note also that there is no violation of energy conditions, since we have only dust matter). Thus, from the above equations, we obtain the following correspondence:

$$-\frac{1}{8\pi G}Q_{\mathscr{D}} \;=\; \varepsilon\dot{\Phi}^2_{\mathscr{D}} - U_{\mathscr{D}} \quad ; \quad -\frac{1}{8\pi G}W_{\mathscr{D}} = 3U_{\mathscr{D}} \;. \tag{11}$$

The correspondence (11) recasts the integrability conditions (5) into a (scale–dependent) Klein–Gordon equation for $\Phi_{\mathscr{D}}$, and $\dot{\Phi}_{\mathscr{D}} \neq 0$:

$$\ddot{\Phi}_{\mathscr{D}} + 3H_{\mathscr{D}}\dot{\Phi}_{\mathscr{D}} + \varepsilon\frac{\partial}{\partial\Phi_{\mathscr{D}}}U(\Phi_{\mathscr{D}},\langle\rho\rangle_{\mathscr{D}}) \;=\; 0 \;. \tag{12}$$

We appreciate that the deviation of the averaged scalar curvature from a constant–curvature model is directly proportional to the potential energy density of the scalar

field. Averaged universe models obeying this set of equations follow, thus, a Friedmannian kinematics with a fundamental matter source, and an effective scalar field source that reflects the shape of spatial hypersurfaces and the shape of their embedding into spacetime. Given the potential in terms of variables of the averaged system, the evolution of these models is fixed (the governing equations are closed). This also potentially fixes coupling parameters, since all involved fields can be traced back to the initial value problem of general relativity.

The morphon formulation of the backreaction problem opens a nice interpretation in terms of energies: a homogeneous model, $Q_{\mathscr{D}} = 0$ (a necessary and sufficient condition to also drop the scale–dependence, if required on every scale), is characterized by the *virial equilibrium condition*:

$$2E^{\mathscr{D}}_{\rm kin} + E^{\mathscr{D}}_{\rm pot} = -\frac{Q_{\mathscr{D}}V_{\mathscr{D}}}{8\pi G} \quad , \quad Q_{\mathscr{D}} = 0 \quad ; \quad E^{\mathscr{D}}_{\rm kin} = \varepsilon\dot{\Phi}^2_{\mathscr{D}}V_{\mathscr{D}} \quad , \quad E^{\mathscr{D}}_{\rm pot} = -U_{\mathscr{D}}V_{\mathscr{D}} \ . \tag{13}$$

Deviations from homogeneity, $Q_{\mathscr{D}} \neq 0$, thus invoke a non–equilibrium dynamics of the morphon in its potential that is dictated by the effective intrinsic curvature of the space in which the fluctuations evolve. Morphon energies are redistributed and can be assigned to the *dark energies*. Dependent on the signs of the backreaction terms (and a sign change may occur by looking at different spatial scales) the morphon can act as a scalar field model for *Dark Matter*, a quintessence model for *Dark Energy*, or it can even play the role of a *classical inflaton*, as we exemplify in the following subsection. (For the different interpretations of scalar fields see the review [25], and for unified views the selection of papers [1, 55, 67, 64], and for scalar Dark Matter e.g. [49, 2, 50].

Example: morphonic inflation

Consider a tube of space time characterized by a gravitational field with no fundamental sources. The $4-$Ricci curvature tensor vanishes everywhere, but not necessarily the $4-$Weyl curvature tensor. Even if this classical vacuum space time is initially foliated into $3-$Ricci flat hypersurfaces, this does not remain so in the dynamical evolution: such an initially prepared homogeneous state is unstable (a fact that we shall explain in the next section), and these hypersurfaces, if perturbed, will necessarily develop into inhomogeneous hypersurfaces featuring non–vanishing averaged curvature invariants, i.e. an intrinsic, on average negative curvature, and a compensating extrinsic curvature due to the embedding of the hypersurfaces into the Ricci–flat space time. Thus, in this picture, the space section will develop a morphonic scalar field that is driven by a Klein–Gordon dynamics and specified by the initial value of its self–interaction potential. While this instability is dynamical, the picture reminds us of the behavior of a fundamental inflaton, where the instability is created by an externally added potential.

We specify initial data according to the analogy of the backreaction variables to the morphon field ($Q^i_{\mathscr{D}} \equiv Q_{\mathscr{D}}(t_i); W^i_{\mathscr{D}} \equiv W_{\mathscr{D}}(t_i)$):

$$U^i_{\mathscr{D}} \equiv -\frac{1}{24\pi G}W^i_{\mathscr{D}} \quad ; \quad \dot{\Phi}^i_{\mathscr{D}} \equiv \sqrt{\frac{-Q^i_{\mathscr{D}}}{8\pi G\varepsilon} + U^i_{\mathscr{D}}} \quad ; \quad \Phi^i_{\mathscr{D}} \equiv \Phi_{\mathscr{D}}(t_i) \, . \tag{14}$$

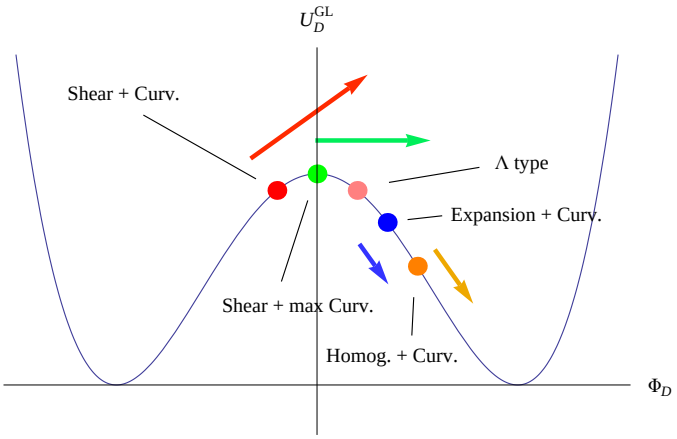


FIGURE 1. The Ginzburg–Landau potential in arbitrary units and the possible initial conditions as well as their physical meaning. All conditions possess some curvature $W^i_{\mathscr{D}} < 0$. The arrows schematically indicate the amplitude of the morphon's initial speed $\dot{\Phi}^i_{\mathscr{D}}$. In the order of the points (from left to right): the first two points dominated by shear fluctuations (red, green) are obtained for $Q^i_{\mathscr{D}} < 0 \Leftrightarrow \dot{\Phi}^{i2}_{\mathscr{D}} > 2(H^{i2}_{\mathscr{D}} + k^i_{\mathscr{D}})$; the next points dominated by expansion fluctuations (blue, pink) for $\dot{\Phi}^{i2}_{\mathscr{D}} < 2(H^{i2}_{\mathscr{D}} + k^i_{\mathscr{D}})$, where the de Sitter–Λ equivalent case has a stiff morphon $\dot{\Phi}^i_{\mathscr{D}} = 0$; the homogeneous case (last point, orange) is obtained for $\dot{\Phi}^{i2}_{\mathscr{D}} = 2(H^{i2}_{\mathscr{D}} + k^i_{\mathscr{D}})$. Figure from [20].

Interestingly, for a homogeneous initial state, $Q_{\mathscr{D}}(t_i) = 0$, the kinetic energy density of the morphon field is initially non–vanishing, and the Klein–Gordon dynamics drives the morphon into a stable fix point, an (assumed) existing minimum of the potential. However, the outcome does not depend much on the initial data for $Q_{\mathscr{D}}$: we could also start with inhomogeneous initial data, e.g. a cosmological constant that is mimicked by a particular morphon, in which case the initial kinetic energy density is zero. Curvature energy is thus converted into kinetic energy, driving the system into an accelerated expansion phase. The value of the potential is necessarily always positive, since the vacuum 3−space has negative intrinsic curvature. The inflationary mechanism is thus the same as the mechanism to create an accelerated expansion in a Quintessence phase.

We realize the inflationary scenario [20] by closing the set of averaged equations with a potential of the generic Ginzburg–Landau form:

$$U^{GL}_{\mathscr{D}} = U_0 \left(\Phi^2_{\mathscr{D}} - \Phi^2_0\right)^2 / \Phi^4_0 \,. \tag{15}$$

The position of the minimum Φ_0 and the amplitude U_0 play different roles: the first one fixes the duration of inflation, while the second sets the size of the Hubble radius at which it happens. This potential has been extensively studied in the context of chaotic inflation [48]. The various initial conditions together with their interpretation in terms of geometrical properties of space are shown in Figure 1.

Combining this purely morphonic picture of inflation created from the Einstein vacuum with a fundamental scalar field, we can establish hybrid inflationary models with two scalar fields, one of them being the morphon that is always present in the case of inhomogeneous universe models.

GLOBAL GRAVITATIONAL INSTABILITY OF THE STANDARD MODEL BACKGROUND

The phase space of exact background states

The space of possible states of an averaged cosmological model, or the space of "generalized backgrounds" has one dimension more than the space of possible homogeneous–isotropic solutions in the standard model framework. This can be seen by introducing adimensional "cosmological parameters". We divide the volume–averaged expansion law by the squared *volume Hubble functional* $H_{\mathscr{D}} := \dot{a}_{\mathscr{D}}/a_{\mathscr{D}}$ introduced before. Then, the expansion law can be expressed as a sum of adimensional average characteristics:

$$\Omega_m^{\mathscr{D}} \, + \, \Omega_\Lambda^{\mathscr{D}} \, + \, \Omega_k^{\mathscr{D}} \, + \, \Omega_W^{\mathscr{D}} \, + \, \Omega_Q^{\mathscr{D}} \, = \, 1 \ , \tag{16}$$

with:

$$\begin{gathered} \Omega_m^{\mathscr{D}} := \tfrac{8\pi G}{3H_{\mathscr{D}}^2}\langle\rho\rangle_{\mathscr{D}} \ ; \ \Omega_\Lambda^{\mathscr{D}} := \tfrac{\Lambda}{3H_{\mathscr{D}}^2} \ ; \ \Omega_k^{\mathscr{D}} := -\tfrac{k_{\mathscr{D}_{\mathrm{i}}}}{a_{\mathscr{D}}^2 H_{\mathscr{D}}^2} \ ; \\ \Omega_W^{\mathscr{D}} := -\tfrac{W_{\mathscr{D}}}{6H_{\mathscr{D}}^2} \ ; \ \Omega_Q^{\mathscr{D}} := -\tfrac{Q_{\mathscr{D}}}{6H_{\mathscr{D}}^2} \ . \end{gathered} \tag{17}$$

Taking for simplicity the constant–curvature parameter and the curvature deviation into a single full curvature parameter, $\Omega_k^{\mathscr{D}} + \Omega_W^{\mathscr{D}} =: \Omega_R^{\mathscr{D}}$, the generalized model offers a *cosmic quartet* of parameters. Furthermore, if we put $\Lambda = 0$, the expansion law defines, for each scale, a two–dimensional phase space of states. A one–dimensional subset of this phase space is formed by "backgrounds" with Friedmannian kinematics (see Figure 2).

We can analyze the fix points and their stability properties in the general dynamical system [16], [65]. The principal outcome of this study is that the standard zero–curvature model forms a *saddle point*; of particular interest are two instability sectors for the standard model, regarded as averaged state: firstly, perturbed homogeneous states are driven into a sector of highly isotropic, negative curvature and accelerated expanding "backgrounds" where backreaction thus mimics Dark Energy behavior over the domain $\mathscr{D}$; secondly, perturbed homogeneous states are driven into a sector of highly anisotropic, positive curvature, collapsing and decelerated "backgrounds" where backreaction thus mimics Dark Matter behavior over the domain $\mathscr{D}$. Concrete models show that the former happens on large scales, and the latter on the scales of galaxy surveys, and also on smaller scales. Thus, qualitatively, the instability sectors identified comply with the aim to trace the dark components back to physical properties, but they also agree with the expected properties of the structure: isotropic, accelerating states on large scales, and highly anisotropic structures on the filamentary distribution of superclusters. Moreover, the curvature properties also meet the expectations: on large scales the Universe is void–dominated and, hence, dominated by negative curvature, while on intermediate scales over–densities are abundant and are characterized by positive curvature.

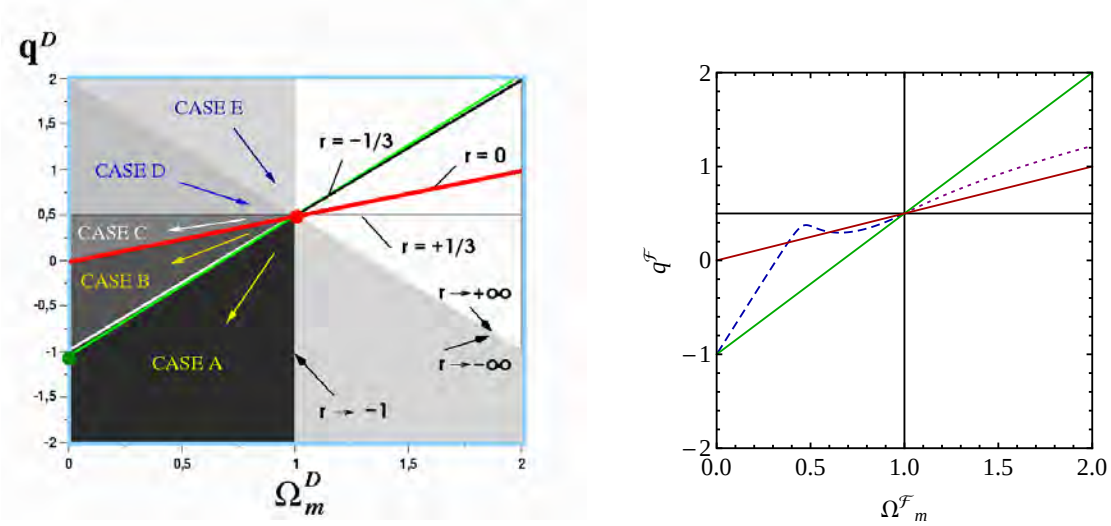


FIGURE 2. Left: "Cosmic phase space" of the solutions of the averaged equations ("generalized backgrounds") in a plane spanned by the volume deceleration parameter $q^{\mathscr{D}} := -\ddot{a}_{\mathscr{D}}/(a_{\mathscr{D}}H^2_{\mathscr{D}}) = 1/2\Omega^{\mathscr{D}}_m + 2\Omega^{\mathscr{D}}_Q - \Omega^{\mathscr{D}}_\Lambda$ and the matter density parameter [16]. It represents a two–dimensional subspace $\{\ \Lambda = 0\ \}$ of the full solution space that would include a cosmological constant. The segments are separated by particular exact scaling solutions of the full problem. We identify the following scaling solutions: all the scaling solutions are represented by straight lines passing through the Einstein–de Sitter model in the center of the diagram $(1/2; 1)$. Models with "Friedmannian" kinematics, but with renormalized parameters form the line $r = 1/3$ (for details see [16], Appendix A). The line $r = 0$ are models with no backreaction on which the parameter $\Omega^{\mathscr{D}}_k$ varies (scale–dependent "Friedmannian models"). Below the line $r = 0$ in the "quintessence phase" we find effective models with subdominant shear fluctuations ($Q_{\mathscr{D}}$ positive, $\Omega^{\mathscr{D}}_Q$ negative). The line $r = -1/3$ mimics a "Friedmannian model" with cosmological constant. The line below $r = -1/3$ in the "phantom quintessence phase" represents the solution inferred from SNLS data (*cf.* [16]), and the point at $(q^{\mathscr{D}}; \Omega^{\mathscr{D}}_m) = (-1.03; 0)$ locates the late–time attractor associated with this solution. Since we have no cosmological constant here, all expanding solutions in the subplane $q^{\mathscr{D}} < 0$ drive the averaged variables away from the standard model featuring a backreaction–driven volume acceleration of effectively isotropic cosmologies that are curvature–dominated at late times.
Right: we show the evolution of phase space orbits (running away from the standard model) for a multiscale model that is explained further below; this model is partitioned into over–dense $\mathscr{M}$ (dotted) and under–dense $\mathscr{E}$ (dashed) regions, their volume fraction being derived from N–body simulations [71]. They are shown here in the same plot for economic reasons and actually live in two different phase spaces corresponding to large scales for the $\mathscr{E}$–regions and small scales for the $\mathscr{M}$–regions. $\mathscr{F}$ denotes one of the regions $\mathscr{M}$ or $\mathscr{E}$.

Dark Energy and Dark Matter hidden in the geometry of space

The fact that the standard model is globally unstable in the phase space of averaged states, and the fact that the instability sectors lie in the right corners to explain Dark Energy and Dark Matter behavior, are both strong qualitative arguments to expect that the conservative explanation of the dark energies through morphon energies is valuable. The underlying mechanism is indeed based on the fundamental existence of the relation between geometrical curvatures and sources dictated by Einstein's equations.

Whether this mechanism is sufficient in a quantitative sense is to date still an open issue. The difficulty to construct quantitative models is to be seen in the need for

non–standard tools, for example perturbation theory on a fixed reference background should be replaced by a fluctuation theory on evolving backgrounds that captures the average over the fluctuations. The question whether perturbations are small can only be answered if we know with respect to which background they are small. Furthermore, since backreaction affects the geometry, it will change the interpretation of observational data, a problem that is intimately related to the generalization of the cosmological model, and to which we shall come below.

Before, we shall in the next section explain the identified mechanism by discussing some physical properties of structure formation and its relation to the interpretation of geometrical curvature invariants and how they mimic the dark sources. We here touch on a deeper problem: backreaction effects account for both, Dark Energy and Dark Matter, simultaneously. Whether, on a given domain, or on an ensemble of domains on a given scale, the morphon mimics Dark Energy or Dark Matter behavior, changes as a function of time and as a function of scale. Moreover, the small–scale contribution to e.g. a Dark Matter behavior requires more sophisticated relativistic models than the dust model used throughout here (e.g. [24], [54]). Considering rotation curves of galaxy halos, missing gravitational sources in clusters or missing sources on cosmological scales always needs different modeling strategies. We try in the following to provide a first step of disentangeling Dark Energy and Dark Matter behavior by explicitly constructing an effective multiscale cosmological model.

Multiscale cosmology and structure–emerging volume acceleration

Contrary to the standard model, where a homogeneous background is used as a standard of reference for the expansion history of the Universe, a background constructed as the average over fluctuating fields introduces a subtle element: while a homogeneous geometry can be characterized locally, an average is nonlocal, since it is determined by the inhomogeneities inside, but also outside the averaging domain, reflecting the nonlocal nature of gravitation. Furthermore, an average incorporates correlations of the local fields. It is this latter which is the key–driver of a repulsiv “effective pressure” that arises in the averaged models.

This “effective pressure” provides the reason why backreaction can produce a volume–accelerating component despite the decelerating nature of the general local acceleration law. This can be seen easily by comparing the local and the volume–averaged Raychaudhuri equation (for vanishing vorticity and pressure that both would also act accelerating on the local level, but only on small scales):

$$\dot{\theta} = \Lambda - 4\pi G\rho + 2\mathrm{II} - \mathrm{I}^2 \quad ; \quad \langle\theta\rangle^{\cdot} = \Lambda - 4\pi G\langle\rho\rangle_{\mathscr{D}} + 2\langle\mathrm{II}\rangle_{\mathscr{D}} - \langle\mathrm{I}\rangle^2_{\mathscr{D}} \ , \qquad (18)$$

where we defined the principal scalar invariants of the expansion tensor Θ_{ij}, $2\mathrm{II} := 2/3\theta^2 - 2\sigma^2$ and $\mathrm{I} := \theta$.

Clearly, by shrinking the domain to a point, both equations agree. However, evaluating the local and averaged invariants,

$$2\mathrm{II} - \mathrm{I}^2 = -\tfrac{1}{3}\theta^2 - 2\sigma^2 \quad ;$$
$$2\left\langle \mathrm{II}\right\rangle_{\mathscr{D}} - \left\langle \mathrm{I}\right\rangle^2_{\mathscr{D}} = \tfrac{2}{3}\left\langle (\theta - \left\langle\theta\right\rangle_{\mathscr{D}})^2\right\rangle_{\mathscr{D}} - 2\left\langle (\sigma - \left\langle\sigma\right\rangle_{\mathscr{D}})^2\right\rangle_{\mathscr{D}} - \tfrac{1}{3}\left\langle\theta\right\rangle^2_{\mathscr{D}} - 2\left\langle\sigma\right\rangle^2_{\mathscr{D}} \quad , \quad (19)$$

gives rise to two additional, positive–definite fluctuation terms, where that for the averaged expansion variance enters with a positive sign. Thus, the time–derivative of a (on some spatial domain $\mathscr{D}$) averaged expansion may be positive despite the fact that the time–derivative of the expansion *at all points* in $\mathscr{D}$ is negative.

In concrete models this variance is the source of a possible large–scale volume–acceleration that would be assigned to Dark Energy in the standard model, while the averaged shear fluctuations mimic an attractive source that would be missing as Dark Matter in the standard model on cosmological scales. Both terms are competing in the kinematical backreaction $Q_{\mathscr{D}}$. Since backreaction depends on scale, it may act in both ways.

We can go one step further and make the scale–dependence explicit by introducing a union of disjoint over–dense regions $\mathscr{M}$ and a union of disjoint under–dense regions $\mathscr{E}$, which both make up the total (homogeneity–scale) region $\mathscr{D}$. The averaged equations can be split accordingly yielding for the kinematical backreaction [17], [71]:

$$Q_{\mathscr{D}} \;=\; \lambda_{\mathscr{M}} Q_{\mathscr{M}} + (1-\lambda_{\mathscr{M}})\, Q_{\mathscr{E}} + 6\lambda_{\mathscr{M}}\,(1-\lambda_{\mathscr{M}})\,(H_{\mathscr{M}} - H_{\mathscr{E}})^2 \;, \qquad (20)$$

where $\lambda_{\mathscr{M}} := |\mathscr{M}|\,/\,|\mathscr{D}|$ denotes the volume–fraction of the over–dense regions compared to the volume of the region $\mathscr{D}$. In a Gaussian random field this fraction would be $0,5$ and would gradually drop in a typical structure formation scenario that clumps matter into small volumes and that features voids that gradually dominate the volume in the course of structure formation.

Ignoring for simplicity the individual backreaction terms on the partitioned domains, the total backreaction features a positive–definite term that describes the variance between the different expansion histories of over– and under–dense regions. It is this term that generates a Dark Energy behavior over the domain $\mathscr{D}$ (see also [59] for a model by Räsänen, and [72, 73, 74, 45] for Wiltshire's model that is based on this term only, but includes a phenomenlogical lapse function to account for different histories in $\mathscr{M}$ and $\mathscr{E}$ regions that, this latter, we cannot implement in the synchronous foliation of a multiscale dust model). If we model non–zero individual backreaction terms by an extrapolation of the leading perturbative mode in second–order perturbation theory [46, 47] that also corresponds to the leading order in a Newtonian non–perturbative model [11], then we even produce a cosmological constant behavior over $\mathscr{D}$, see Figure 3. In other words, the fact that, physically, over–dense regions tend to be gravitationally bound, i.e. do not partake significantly in the global expansion, already produces a large–scale "kinematical pressure" as a source of volume acceleration. A homogeneous background simply cannot account for this difference.

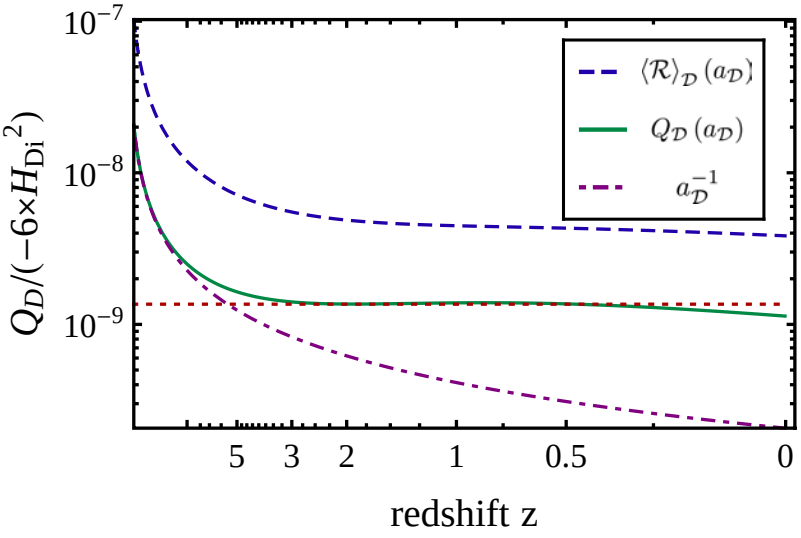


FIGURE 3. Plot of the evolution of $Q_{\mathscr{D}}$ and $\langle R\rangle_{\mathscr{D}}$ in terms of the global scale factor $a_{\mathscr{D}}$. For comparison a line with a simple $a_{\mathscr{D}}^{-1}$–scaling is added, that corresponds to the leading mode in second–order perturbation theory, and one that is constant. $Q_{\mathscr{D}}$ and $\langle R\rangle_{\mathscr{D}}$ are normalized by $-6H^2_{\mathscr{D}_{\rm i}}$ so that the values at the initial time represent $\Omega_Q^{\mathscr{D}_{\rm i}}$ and $\Omega_R^{\mathscr{D}_{\rm i}}$. We appreciate that the backreaction terms feature an approximate cosmological constant behavior on the homogeneity scale despite the assumption of $a_{\mathscr{F}}^{-1}$–scaling on the partitioned domains $\mathscr{F}=\mathscr{M},\mathscr{E}$. Physically, this result can be attributed to the expansion variance between the subdomains and, hence, this latter is identified as the key effect to produce a global Dark Energy–like behavior of the backreaction terms. Figure from [71].

INHOMOGENEOUS AND EFFECTIVE METRICS

Some notes on relativistic perturbation theories

Consider a spatial metric form $\mathbf{g}$ with coefficients g_{ij} in an exact (co–tangential) basis $\mathbf{d}X^i\otimes\mathbf{d}X^j$. We can write any metric as a quadratic form of deformation one–forms, $\mathbf{g}=\delta_{ab}\eta^a\otimes\eta^b$, i.e. in terms of coefficients, $g_{ij}=\delta_{ab}\eta^a_{\ i}\eta^b_{\ j}$. Now, such a metric form is *homogeneous*, i.e. its Ricci tensor vanishes everywhere, if there exist functions f^a, such that the one–forms can be written as exact forms, $\eta^a\equiv\mathbf{d}f^a$. In other words, if we can find a coordinate transformation $x^i=f^{a\equiv i}(X^j,t)$ that transforms the Euclidean metric coefficients in a new basis, $\mathbf{d}x^i\otimes\mathbf{d}x^j$, $\delta_{ij}dx^idx^j=\delta_{ab}f^a_{\ |i}f^b_{\ |j}dX^idX^j$, with a vertical slash denoting partial spatial derivative, into the metric coefficients g_{ij}, then these latter are just a rewriting of the homogeneous space. Given this remark, any perturbation theory that features metric forms of the integrable form, does not describe relativistic inhomogeneities; metric coefficients of the form $g_{ij}=\delta_{ab}f^a_{\ |i}f^b_{\ |j}$ describe Newtonian (Lagrangian) perturbations on a flat background space. A truely relativistic perturbation theory deforms the background geometry, in other words, the perturbations live in a perturbed space, not on a reference background. This remark also shows that relativistic perturbation terms can never contain full divergences, since this latter needs integrable

one–form fields.

In light of these introductory remarks, an inhomogeneous relativistic metric produces curvature that, if volume–averaged on some domain, does not obey a conservation law (as can be explicitly seen in the coupling equation to the fluctuations (5)) in the sense that it would always average out to zero; for details on curvature estimates see [17]). This fact in itself shows the existence of a dynamical evolution of an averaged curvature, as soon as structures form. On the contrary, standard perturbation theory formulated on a fixed flat background is constructed such that the averages always vanish on the background, demonstrating the limited nature of results obtained by standard perturbative arguments.

Another perturbative argument aims to justify the validity of the homogeneous geometry, even down to the scales of neutron stars [37], since perturbations of the metric remain small with respect to the flat background. This argument does not contradict the existence of a large backreaction effect, since these latter depend on first and second derivatives of the metric [19], [40], [62]. Also, the perturbations are considered on a flat background that does not interact with structure. As we explained in detail, the perturbations may be small on a different (physical) background, in which case a perturbation may already live in a background with strong curvature (a zero–order effect). It is therefore not fruitful to argue against the relevance of backreaction within standard limited schemes, but rather an effort to generalize perturbation theory is needed.

Template metrics and effective distances

For the construction of an effective cosmological evolution model, as outlined above, a metric needs not be specified. The need for the construction of an effective metric in these models arises, since measured redshifts have to be interpreted in terms of distances along the light cone. Given an explicit, generic and realistic, inhomogeneous metric, the need for the construction of effective metrics does not arise. Also, if we succeed to understand the evolution of light cone averages in relation to distances, then also here an explicit metric will not be needed [60, 61], [21].

The idea of an effective cosmological metric comes from the “fitting problem”, that has been particularly emphasized by George Ellis already in the 70’s [26]. The observation was that an inhomogeneous metric does not average out to a homogeneous metric that forms a solution of general relativity. Not only the nonlinearity of the theory, but also simple arguments of a non–commutativity [28] between evolution equations and the averaging operation, give rise to the need to find a “best–fit”, we may call it “template” geometry, that inherits homogeneity and (almost–)isotropy on the large scales and, at the same time, incorporates the inhomogeneous structure “on average” (see also the early practical implementations of this problem [31, 32], [27], [34]).

For the solution of the *fitting problem* various strategies have been proposed (see [28] and references therein). One strategy, that allows to explicitly perform a “smoothing” of an inhomogeneous metric into a constant–curvature metric at one instant of time, is based on Ricci–flow theory: one notices that a smoothing operation of metrical properties can be put into practice by a *rescaling* of the metric in the direction of its Ricci curvature. The scaling equations for realizing this are well–studied, and the

rescaling flow results in a constant–curvature metric that carries "dressed" cosmological variables [12], [13]. These incorporate intrinsic curvature backreaction terms describing the difference to the "bare" cosmological parameters as they are obtained through kinematical averaging.

Reinterpretation of observational data

The standard method of interpreting observations is to construct the light cone $ds^2 = 0$ from the line–element $ds^2 = -dt^2 + g^{\mathrm{hom}}_{ij} dX^i dX^j$, where the coefficients g^{hom}_{ij} are given in the form of a constant–curvature (FLRW) metric, and then to calculate the luminosity distance $d_L(z)$ in this metric for a given observed redshift z. Assuming this metric for the inhomogeneous Universe implies the conjecture that the FLRW metric is the correct "template" of an effective cosmological metric. However, the integrated exact equations (the integral properties of a general inhomogeneous model) are not compatible with this metric, simply because the averaged curvature is assumed to be of the form $\langle R \rangle_{\mathscr{D}} = 6ka^{-2}$ on all scales. Improving the metric template slightly, by replacing the global scale factor $a(t)$ through the volume–scale factor $a_{\mathscr{D}}(t)$ and the integration constant k through the domain–dependent integration constant $k_{\mathscr{D}}$, renders this metric implicitly scale–dependent [53]. As we explained, this is not enough since the averaged curvature couples to the inhomogeneities and in general deviates from the $a_{\mathscr{D}}^{-2}$–behavior. What we can do as a first approximation, and this would render the metric compatible with the kinematical average properties, is to introduce the exact averaged curvature in place of the constant curvature in this metric form [43].

The resulting effective space time metric consists of a synchronous foliation of constant–curvature metrics that are, however, parametrized by the exact integral properties of the inhomogeneous curvature, thus they "repair" the standard template metric as for the evolution properties of spatial variables. Such a construction can be motivated by Ricci–flow smoothing, that guarantees the existence of smoothed–out constant curvature sections at one instant of time, and by assuming that the intrinsic backreaction terms are subdominant, so that we can parametrize the metric by "bare" kinematical averages. To stack these hypersurfaces together introduces, however, an inhomogeneous light cone structure [51], [58]. Ideally, one would wish to smooth the light cone too, which is also possible by employing Ricci flow techniques. Improving this first approach to a template metric is needed and this is work in progress.

The result of employing the improved template metric described above is a change in the luminosity distance that would alter the interpretation of all observational data formerly based on FLRW distances. Examples for the multiscale models investigated in [71] are presented in Figure 4. Although this investigation certainly needs refinement, we already appreciate a signature of the different curvature evolution that furnishes a clearcut prediction for future observations (see [43] for details).

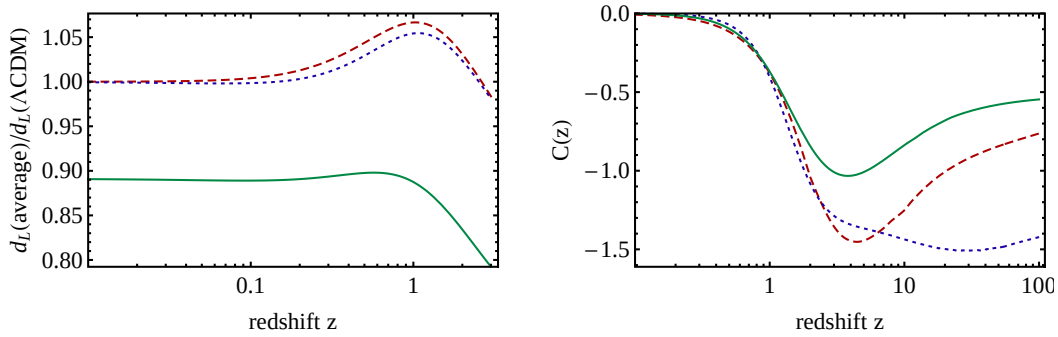


FIGURE 4. Left: Comparison of the luminosity distances of the multiscale models investigated in [71] and based on the template metric of [43], with the one of a flat ΛCDM model with $h = 0.7$ and $\Omega_m = 0.27$. On top a model where we force the volume scale factor $a_{\mathscr{D}}$ to follow the ΛCDM evolution. Despite this assumption, the changing curvature affects the luminosity distance. The luminosity distances in these models show a significant feature at a redshift of around 1, when compared with the best fit ΛCDM model, which may be looked for in the SN data. The curve below is a model with $a_{\mathscr{D}}^{-1}$–scaling. For comparison we also included the luminosity distance of the best fit model of [43]. Because of a different Hubble rate of $h = 0.7854$ it lies below the others from the beginning. This model does not significantly show the distinct feature of the other two models around a redshift of 1, due to the assumption of a single–scale cosmology.
Right: Values of Clarkson's C–function [23] for the best fit model of [43] (top), the model where the scale–factor is forced to follow the ΛCDM evolution (middle), and the model with $a_{\mathscr{D}}^{-1}$–scaling (bottom). Recall that, for every Friedmann model, $C(z)$ vanishes exactly on all scales and for all redshifts. For the inhomogeneous models shown in the plot, this function has a minimum which may serve as observational evidence for the effective cosmologies, as proposed in [43]. As both multiscale models show, it is not even necessary to measure derivatives of distance, since the feature is already present in the distance.
Figure from [71].

Acknowledgements:

Thanks go to my collaborators with whome I share some of the presented results. It is a pleasure to thank the organizers of the workshop for inviting this contribution. This work is supported by "Fédération de Physique André–Marie Ampère" of Université Lyon 1 and École Normale Supérieure de Lyon.

REFERENCES

1. A. Arbey: Dark fluid: A complex scalar field to unify dark energy and dark matter. *Phys. Rev. D* **74**, 043516 (2006).
2. A. Bernal, T. Matos and D. Nunez: Flat Central Density Profiles from Scalar Field Dark Matter Halo. (Rev. Mex. AA Publ.) arXiv:astro-ph/0303455 (2003).
3. T. Biswas, A. Notari and W. Valkenburg: Testing the Void against Cosmological data: fitting CMB, BAO, SN and H0. arXiv:1007.3065 (2010).
4. K. Bolejko and L. Andersson: Apparent and average acceleration of the Universe. *JCAP* **10**, 003 (2008).
5. I.A. Brown, G. Robbers and J. Behrend: Averaging Robertson-Walker cosmologies. *JCAP* **4**, 016 (2009).
6. I.A. Brown, J. Behrend and K.A. Malik: Gauges and Cosmological Backreaction. *JCAP* **11**, 027 (2009).
7. T. Buchert and J. Ehlers: Averaging inhomogeneous Newtonian cosmologies. *Astron. Astrophys.* **320**, 1 (1997).
8. T. Buchert: On average properties of inhomogeneous cosmologies. In: 9th JGRG Meeting, Hiroshima 1999, Y. Eriguchi et al. (eds.), *J.G.R.G.* **9**, 306–321 (2000), arXiv:gr–qc/0001056
9. T. Buchert: On average properties of inhomogeneous fluids in general relativity: 1. dust cosmologies. *Gen. Rel. Grav.* **32**, 105 (2000).
10. T. Buchert: On average properties of inhomogeneous fluids in general relativity: 2. perfect fluid cosmologies. *Gen. Rel. Grav.* **33**, 1381 (2001).
11. T. Buchert, M. Kerscher and C. Sicka: Backreaction of inhomogeneities on the expansion: The evolution of cosmological parameters. *Phys. Rev. D* **62**, 043525 (2000).
12. T. Buchert and M. Carfora: Regional averaging and scaling in relativistic cosmology. *Class. Quant. Grav.* **19**, 6109 (2002).
13. T. Buchert and M. Carfora: Cosmological parameters are 'dressed'. *Phys. Rev. Lett.* **90**, 031101-1-4 (2003).
14. T. Buchert: A cosmic equation of state for the inhomogeneous Universe: can a global far–from–equilibrium state explain Dark Energy? *Class. Quant. Grav.* **22**, L113 (2005).
15. T. Buchert: On globally static and stationary cosmologies with or without a cosmological constant and the Dark Energy problem. *Class. Quant. Grav.* **23**, 817 (2006).
16. T. Buchert, J. Larena and J.–M. Alimi: Correspondence between kinematical backreaction and scalar field cosmologies – the 'morphon field'. *Class. Quant. Grav.* **23**, 6379 (2006).
17. T. Buchert and M. Carfora: On the curvature of the present day universe. *Class. Quant. Grav.* **25**, 195001 (2008).
18. T. Buchert: Dark Energy from structure – a status report. *Gen. Rel. Grav.* **40**, 467 (2008).
19. T. Buchert, G.F.R. Ellis and H. van Elst: Geometrical order–of–magnitude estimates for spatial curvature in realistic models of the Universe. *Gen. Rel. Grav.* **41**, 2017 (2009).
20. T. Buchert and N. Obadia: Effective inhomogeneous inflation I: curvature inhomogeneities of the Einstein vacuum. arXiv:1010.4512 (2010).
21. T. Buchert, H. van Elst and D.J. Schwarz: The averaging problem on the lightcone in inhomogeneous cosmology. *in preparation.*
22. M.–N. Célérier: Inhomogeneities in the Universe with exact solutions of General Relativity. Proceedings of the Invisible Universe International Conference, held in Paris, 29 June – 3 July 2009, eds. J.–M. Alimi, A. Füzfa and P.–S. Corasaniti, *AIP Conference Proceedings: Volume* **1241**, pp. 767–775 (2010).
23. C. Clarkson, B. Bassett and T. H.-C. Lu: A General Test of the Copernican Principle. *Phys. Rev. Lett.* **101**, 011301 (2008).
24. F. I. Cooperstock and S. Tieu: Galactic dynamics via general relativity: a compilation and new developments. *Int. J. Mod. Phys. A* **22**, 2293 (2007).
25. E.J. Copeland, M. Sami and S. Tsujikawa: Dynamics of Dark Energy. *Int. J. of M. Phys. D* **15**, 1753 (2006).
26. G.F.R. Ellis: Relativistic cosmology: its nature, aims and problems. In *General Relativity and Gravitation* (D. Reidel Publishing Company, Dordrecht, 1984), pp. 215–288 (1984).

27. G.F.R. Ellis and W. Stoeger: The 'fitting problem' in cosmology. *Class. Quant. Grav.* **4**, 1697 (1987).
28. G.F.R. Ellis and T. Buchert: The Universe seen at different scales. *Phys. Lett. A.* (Einstein Special Issue) **347**, 38 (2005).
29. K. Enqvist: Lemaitre Tolman Bondi model and accelerating expansion. *Gen. Rel. Grav.* **40**, 451 (2008).
30. R.J. Foley et al.: Spectroscopy of High–Redshift Supernovae from the Essence Project: The First Four Years. *AJ* **137** 3731 (2009).
31. T. Futamase: An approximation scheme for constructing inhomogeneous universes in general relativity. *Mon. Not. Roy. Astron. Soc.* **237**, 187 (1989).
32. T. Futamase: Averaging of a locally inhomogeneous realistic universe. *Phys. Rev. D* **53**, 681 (1996).
33. M. Gasperini, G. Marozzi and G. Veneziano: A covariant and gauge–invariant formulation of the cosmological "backreaction". *JCAP* **2**, 009 (2010).
34. C. Hellaby: Volume matching in Tolman models. *Gen. Rel. Grav.* **20**, 1203 (1988).
35. M. Hicken et al.: Improved Dark Energy Constraints from ˜100 New CfA Supernova Type Ia Light Curves. *Astrophys. J.* **700** 1097 (2009).
36. P. Hunt and S. Sarkar: Constraints on large scale inhomogeneities from WMAP–5 and SDSS: confrontation with recent observations. *Mon. Not. Roy. Astron. Soc.* **401**, 547 (2010).
37. A. Ishibashi and R. M. Wald: Can the acceleration of our universe be explained by the effects of inhomogeneities? *Class. Quant. Grav.* **23**, 235 (2006).
38. Kerscher, M., Buchert, T., Futamase, T.: On the abundance of collapsed objects. *Astrophys. J.* **558**, L79 (2001).
39. E.W. Kolb, S. Matarrese, A. Notari and A. Riotto: Effect of inhomogeneities on the expansion rate of the Universe. *Phys. Rev. D* **71**, 023524 (2005).
40. E.W. Kolb, V. Marra and S. Matarrese: Description of our cosmological spacetime as a perturbed conformal Newtonian metric and implications for the backreaction proposal for the accelerating Universe. *Phys. Rev. D* **78** 103002 (2008).
41. E.W. Kolb, V. Marra and S. Matarrese: Cosmological background solutions and cosmological backreactions. *Gen. Rel. Grav.* **42** 1399 (2010).
42. M. Kowalski et al.: Improved Cosmological Constraints from New, Old, and Combined Supernova Data Sets. *Astrophys. J.* **686** 749 (2008).
43. J. Larena, J.–M. Alimi, T. Buchert, M. Kunz and P.–S. Corasaniti: Testing backreaction effects with observations. *Phys. Rev. D* **79**, 083011 (2009).
44. J. Larena: Spatially averaged cosmology in an arbitrary coordinate system. *Phys. Rev. D* **79**, 084006 (2009).
45. B.M. Leith, S.C.C. Ng and D.L. Wiltshire: Gravitational Energy as Dark Energy: Concordance of Cosmological Tests. *Astrophys. J. Lett.* **672**, L91 (2008).
46. N. Li and D.J. Schwarz: On the onset of cosmological backreaction. *Phys. Rev. D* **76**, 083011 (2007).
47. N. Li and D.J. Schwarz: Scale dependence of cosmological backreaction. *Phys. Rev. D* **78**, 083531 (2008).
48. A.D. Linde: Chaotic Inflation. *Phys. Lett. B* **129**, 177 (1983).
49. T. Matos and L.A. Ureña-López: Quintessence and scalar dark matter in the Universe. *Class. Quant. Grav.* **17**, L75 (2001).
50. T. Matos et al.: Dynamics of Scalar Field Dark Matter With a Cosh–like Potential. *Phys. Rev. D* **80**, 123521 (2009).
51. L. Mersini–Houghton and E. Kafexhiu: Nontrivial geometries: Bounds on the curvature of the universe. *Astroparticle Physics* **29**, 167 (2008).
52. A. Paranjape and T.P. Singh: The possibility of cosmic acceleration via spatial averaging in Lemaître Tolman Bondi models. *Class. Quant. Grav.* **23**, 6955 (2006).
53. A. Paranjape and T.P. Singh: Explicit cosmological coarse graining via spatial averaging. *Gen. Rel. Grav.* **40**, 139 (2008).
54. L. Rizzi, S.L. Cacciatori, V. Gorini, A. Kamenshchik and O.F. Piattella: Dark matter effects in vacuum spacetime. *Phys. Rev. D*, **82**, 027301 (2010).

55. A. Pérez-Lorenzana, M. Montesinos and T. Matos: Unification of cosmological scalar fields. *Phys. Rev. D* **77**, 063507 (2008).
56. S. Räsänen: Dark energy from backreaction. *JCAP* **2**, 003 (2004).
57. S. Räsänen: Accelerated expansion from structure formation. *JCAP* **11**, 003 (2006).
58. S. Räsänen: Comment on 'Nontrivial geometries: bounds on the curvature of the Universe'. *Astroparticle Physics* **30**, 216 (2008).
59. S. Räsänen: Evaluating backreaction with the peak model of structure formation. *JCAP* **4**, 026 (2008).
60. S. Räsänen: Light propagation in statistically homogeneous and isotropic dust universes. *JCAP* **2**, 011 (2009).
61. S. Räsänen: Light propagation in statistically homogeneous and isotropic universes with general matter content. *JCAP* **3**, 018 (2010).
62. S. Räsänen: Applicability of the linearly perturbed FRW metric and Newtonian cosmology. *Phys. Rev. D* **81**, 103512 (2010).
63. M. Roos: Dark Matter: The evidence from astronomy, astrophysics and cosmology. arXiv:1001.0316 (2010).
64. X. Roy and T. Buchert: Chaplygin gas and effective description of inhomogeneous universe models in general relativity. *Class. Quant. Grav.* **27**, 175013 (2010).
65. X. Roy, S. Carloni, T. Buchert and N. Obadia: Global gravitational instability of FLRW backgrounds I: dynamical system analysis of the dark sectors. *In preparation*.
66. P. Ruiz–Lapuente: Dark energy, gravitation and supernovae. *Class. Quant. Grav.* **24**, R91 (2007).
67. V. Sahni and L. Wang: New cosmological model of quintessence and dark matter. *Phys. Rev. D* **62**, 103517 (2000).
68. M. Seikel and D.J. Schwarz: Model– and calibration–independent test of cosmic acceleration. *JCAP* **2**, 024 (2009).
69. F. Sylos Labini and Y.V. Baryshev: Testing the Copernican and Cosmological Principles in the local universe with galaxy surveys. *JCAP* **6**, 021 (2010).
70. M. Vonlanthen, S. Räsänen and R. Durrer: Model–independent cosmological constraints from the CMB. *JCAP* **8** 023 (2010).
71. A. Wiegand and T. Buchert: Multiscale cosmology and structure–emerging Dark Energy: a plausibility analysis. *Phys. Rev. D.* **82**, 023523 (2010).
72. D.L. Wiltshire: Cosmic clocks, cosmic variance and cosmic averages. *New Journal of Physics* **9**, 377 (2007).
73. D.L. Wiltshire: Exact Solution to the Averaging Problem in Cosmology. *Phys. Rev. Lett.* **25**, 251101 (2007).
74. D.L. Wiltshire: Average observational quantities in the timescape cosmology. *Phys. Rev. D* **80**, 123512 (2009).

Gravitational energy as dark energy: cosmic structure and apparent acceleration[1]

David L. Wiltshire

Department of Physics & Astronomy, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand

Abstract. Below scales of about $100\,h^{-1}$ Mpc our universe displays a complex inhomogeneous structure dominated by voids, with clusters of galaxies in sheets and filaments. The coincidence that cosmic expansion appears to start accelerating at the epoch when such structures form has prompted a number of researchers to question whether dark energy is a signature of a failure of the standard cosmology to properly account, on average, for the distribution of matter we observe. Here I discuss the timescape scenario, in which cosmic acceleration is understood as an apparent effect, due to gravitational energy gradients that grow when spatial curvature gradients become significant with the nonlinear growth of cosmic structure. I discuss conceptual issues related to the averaging problem, and their impact on the calibration of local geometry to the solutions of the volume–average evolution equations corrected by backreaction, and the question of nonbaryonic dark matter in the timescape framework. I further discuss recent work on defining observational tests for average geometric quantities which can distinguish the timescape model from a cosmological constant or other models of dark energy.

Keywords: dark energy, theoretical cosmology, observational cosmology
PACS: 98.80.-k 98.80.Es 95.36.+x 98.80.Jk

INTRODUCTION

This conference is titled a "Conference on Two Cosmological Models" but I think that what have been presented are certainly more than two cosmological models. From the point of view of proponents of the standard cosmology, the conference might seem to be dealing with too many cosmological models. The fact that there are a lot of ideas on the table is natural at any point in the history of science when observations present a fundamental crisis. We have reached such a point, given that our current standard cosmology only works by invoking unknown sources of "dark energy" and "dark matter", which supposedly make up most of the stuff in the universe.

All of the models presented, including the standard ΛCDM cosmology, could be said to be relativistic in the sense that they obey either Einstein's equations or some extension of Einstein gravity with a geometric diffeomorphism invariant action. What is at issue, however, is the manner in which different models seek to explain the observed large scale structure and motion of objects in the universe. Do we add new fields or modifications to the gravitational action, whose only influence is on cosmological scales, or do we seek to find deeper answers in the principles of general relativity?

---

[1] Based on a presentation at the *International Conference on Two Cosmological Models*, Universidad Iberoamericana, Mexico City, 17-19 November, 2010; to appear in the Proceedings, ed. J. Auping.

These are questions that Einstein struggled with when he first applied general relativity to cosmology [1]. He thought of spacetime as being a relational structure, and therefore the introduction of a cosmological constant – a vacuum energy in the fabric of space which made no direct connection to inertial properties of matter – was not a step he took lightly. I will take the viewpoint that rather than adding further epicycles to the gravitational action, the cosmological observations which we currently interpret in terms of dark energy are inviting us to think more deeply about the foundations of general relativity. There are questions in general relativity – relating to coarse–graining, averaging and the definition of energy in such contexts – which have never been fully resolved. These are the questions which I believe are of relevance to cosmology.

In this paper I will review the conceptual basis [2, 3, 4] and observational tests [5] of a cosmology model [2, 6], which represents a new approach to understanding the phenomenology of dark energy as a consequence of the effect of the growth of inhomogeneous structures. The basic idea, outlined in a nontechnical manner in ref. [7], is that as inhomogeneities grow one must consider not only their backreaction on average cosmic evolution, but also the variance in the geometry as it affects the calibration of clocks and rulers of ideal observers. Dark energy is then effectively realised as a misidentification of gravitational energy gradients.

Although the standard Lambda Cold Dark Matter (ΛCDM) model provides a good fit to many tests, there are tensions between some tests, and also a number of puzzles and anomalies. Furthermore, at the present epoch the observed universe is only statistically homogeneous once one samples on scales of 150–300 Mpc. Below such scales it displays a web–like structure, dominated in volume by voids. Some 40%–50% of the volume of the present epoch universe is in voids with $\delta\rho/\rho\sim -1$ on scales of $30h^{-1}$ Mpc [8], where h is the dimensionless parameter related to the Hubble constant by $H_0 = 100h\,\mathrm{km\ sec^{-1}\,Mpc^{-1}}$. Once one also accounts for numerous minivoids, and perhaps also a few larger voids, then it appears that the present epoch universe is void-dominated. Clusters of galaxies are spread in sheets that surround these voids, and in thin filaments that thread them.

A number of different approaches have been taken to study inhomogeneous cosmologies. One large area of research is that of exact solutions of Einstein's equations (see, e.g., ref. [9]), and of the Lemaître–Tolman–Bondi [10] (LTB) dust solution in particular. While one may mimic any luminosity distance relation with LTB models, generally the inhomogeneities required to match type Ia supernovae (SneIa) data are much larger than the typical scales of voids described above. Furthermore, one must assume the unlikely symmetry of a spherically symmetric universe about our point, which violates the Copernican principle. It is my view that while the LTB solutions are interesting toy models, one should retain the Copernican principle in a statistical sense, and one should seriously try to model the universe with those scales of inhomogeneity that we actually observe.

One particular consequence of a matter distribution that is only statistically homogeneous, rather than exactly homogeneous, is that when the Einstein equations are averaged they do not evolve as a smooth Friedmann–Lemaître–Robertson–Walker (FLRW) geometry. Instead the Friedmann equations are supplemented by additional backreaction

terms[2] [12]. Whether or not one can fully explain the expansion history of the universe as a consequence of the growth of inhomogeneities and backreaction, without a fluid–like dark energy, is the subject of ongoing debate [13].

A typical line of reasoning against backreaction is that of a plausibility argument [14]: if we *assume* a FLRW geometry with small perturbations, and estimate the magnitude of the perturbations from the typical rotational and peculiar velocities of galaxies, then the corrections of inhomogeneities are consistently small. This would be a powerful argument, were it not for the fact that at the present epoch galaxies are not homogeneously distributed. The Hubble Deep Field reveals that galaxy clusters were close to being homogeneous distributed at early epochs, but following the growth voids at redshifts $z \lesssim 1$ that is no longer the case today. Therefore galaxies cannot be consistently treated as randomly distributed gas particles on the $30h^{-1}$ Mpc scales [8] that dominate present cosmic structure below the scale of statistical homogeneity.

Over the past few years I have developed a new physical interpretation of cosmological solutions within the Buchert averaging scheme [2, 3, 6]. I start by noting that in the presence of strong spatial curvature gradients, not only should the average evolution equations be replaced by equations with terms involving backreaction, but the physical interpretation of average quantities must also account for the differences between the local geometry and the average geometry. In other words, geometric variance can be just as important as geometric averaging when it comes to the physical interpretation of the expansion history of the universe.

I proceed from the fact that structure formation provides a natural division of scales in the observed universe. As observers in galaxies, we and the objects we observe in other galaxies are necessarily in bound structures, which formed from density perturbations that were greater than critical density. If we consider the evidence of the large scale structure surveys on the other hand, then the average location by volume in the present epoch universe is in a void, which is negatively curved. We can expect systematic differences in spatial curvature between the average mass environment, in bound structures, and the volume-average environment, in voids.

Spatial curvature gradients will in general give rise to gravitational energy gradients, and herein lie the issues which I believe are key to understanding the phenomenon of dark energy. The definition of gravitational energy in general relativity is notoriously subtle. This is due to the equivalence principle, which means that we can always get rid of gravity near a point. As a consequence, the energy, momentum and angular momentum associated with the gravitational field, which have macroscopic effects on the relative calibrations of the clocks and rulers of observers, cannot be described by local quantities encoded in a fluid-like energy-momentum tensor. Instead they are at best *quasi-local* [15]. There is no general agreement on how to deal with quasi-local gravitational energy. It is my view that since the issue has its origin in the equivalence principle, we must return to first principles and reconsider the equivalence principle in the context of cosmological averages.

---

[2] For a general review of averaging and backreaction see, e.g., ref. [11].

THE COSMOLOGICAL EQUIVALENCE PRINCIPLE

In laying the foundations of general relativity, Einstein sought to refine our physical understanding of that most central physical concept: *inertia*. As he stated: "In a consistent theory of relativity there can be be no inertia relatively to 'space', but only an inertia of masses relatively to one another" [1]. This is the general philosophy that underlies Mach's principle, which strongly guided Einstein. However, the refinement of the understanding of inertia that Einstein left us with in relation to gravity, the Strong Equivalence Principle (SEP), only goes part-way in addressing Mach's principle.

Mach's principle may be stated [16, 17]: *"Local inertial frames (LIFs) are determined through the distributions of energy and momentum in the universe by some weighted average of the apparent motions"*. The SEP says nothing about the average effect of gravity, and therefore nothing about the suitable "weighted average of the apparent motions" of the matter in the universe. Since gravity for ordinary matter fields obeying the strong energy condition is universally attractive, the spacetime geometry of a universe containing matter is not stable, but is necessarily dynamically evolving. Therefore, accounting for the average effect of matter to address Mach's principle means that any relevant frame in cosmological averages is one in which time symmetries of the Lorentz group in LIFs are removed.

My proposal for applying the equivalence principle on cosmological scales is to deal with the average effects of the evolving density by extending the SEP to larger regional frames while removing the time translation and boost symmetries of the LIF to define a *Cosmological Equivalence Principle* as follows [3]:

At any event, always and everywhere, it is possible to choose a suitably defined spacetime neighbourhood, the cosmological inertial frame (CIF), in which average motions (timelike and null) can be described by geodesics in a geometry that is Minkowski up to some time-dependent conformal transformation,

$$\mathrm{d}s^2_{\mathrm{CIF}} = a^2(\eta)\left[-\mathrm{d}\eta^2 + \mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2)\right]. \tag{1}$$

Since the average geometry is a time–dependent conformal scaling of Minkowski space, the CEP reduces to the standard SEP if $a(\eta)$ is constant, or alternatively over very short time intervals during which the time variation of $a(\eta)$ can be neglected. The relation to cosmological averages is understood by the fact that (1) is the spatially flat FLRW metric. In the standard cosmology this is taken to be the geometry of the whole universe. Here, however, the whole universe is inhomogeneous while its geometry is restricted by the requirement that it is possible to always choose (1) as a regional average. This would rule out geometries with global anisotropies, such as Bianchi models, while hopefully leaving enough room to describe an inhomogeneous but statistically homogeneous universe like the one we observe.

To understand why an average geometry (1) is a relevant average reference geometry for the relative calibration of rulers and clocks in the absence of global Killing vectors, let us construct what I will call the *semi-tethered lattice* by the following means. Take a lattice of observers in Minkowski space, initially moving isotropically away from each nearest neighbour at uniform initial velocities. The lattice of observers are chosen to be equidistant along mutual oriented $\hat{x}$, $\hat{y}$ and $\hat{z}$ axes. Now suppose that the observers are

each connected to six others by tethers of negligible mass and identical tension along the mutually oriented spatial axes. The tethers are not fixed but unwind freely from spools on which an arbitrarily long supply of tether is wound. The tethers initially unreel at the same uniform rate, representing a "recession velocity". Each observer carries synchronised clocks, and at a prearranged local proper time all observers apply brakes to each spool, the braking mechanisms having been pre-programmed to deliver the same impulse as a function of local time.

The semi-tethered lattice experiment is directly analogous to the decelerating volume expansion of (1) due to some average homogeneous matter density, because it maintains the homogeneity and isotropy of space over a region as large as the lattice. Work is done in applying the brakes, and energy can be extracted from this – just as kinetic energy of expansion of the universe is converted to other forms by gravitational collapse. Since brakes are applied in unison, however, there is *no net force on any observer in the lattice*, justifying the *inertial frame* interpretation, even though each observer has a nonzero 4-acceleration with respect to the global Minkowski frame. The braking function may have an arbitrary time profile; provided it is applied uniformly at every lattice site the clocks will remain synchronous in the comoving sense, as all observers have undergone the same relative deceleration.

Whereas the Strong Equivalence Principle leads us to define local inertial frames, related to each other by local Lorentz transformations acting at a point, the Cosmological Equivalence Principle refers to a *collective* symmetry of the background. In defining the averaging region of the CIF we are isolating just that part of the volume expansion which is regionally homogeneous and isotropic, and which is determined by the regionally homogeneous part of the background density.

Let us now consider two sets of disjoint semi-tethered lattices, with identical initial local expansion velocities, in a background static Minkowski space. (See Fig. 1(a).) Observers in the first congruence apply brakes in unison to decelerate homogeneously and isotropically at one rate. Observers in the second congruence do so similarly, but at a different rate. Suppose that when transformed to a global Minkowski frame, with time t, that at each time step the magnitudes of the 4–decelerations satisfy $\alpha_1(t) > \alpha_2(t)$ for the respective congruences. By special relativity, since members of the first congruence decelerate more than those of the second congruence, at any time t their proper times satisfy $\tau_1 < \tau_2$. The members of the first congruence age less quickly than members of the second congruence.

By the CEP, the case of volume expansion of two disjoint regions of different average density in the actual universe is entirely analogous. The equivalence of the circumstance rests on the fact that the expansion of the universe was extremely uniform at the time of last scattering, by the evidence of the CMB. At that epoch all regions had almost the *same* density – with tiny fluctuations – and the same uniform Hubble flow. At late epochs, suppose that in the frame of any average cosmological observer there are expanding regions of *different* density which have decelerated by different amounts by a given time, t, according to that observer. Then by the CEP the local proper time of the comoving observers in the denser region, which has decelerated more, will be less than that of the equivalent observers in the less dense region which has decelerated less. (See Fig. 1(b).) Consequently the *proper time of the observers in the more dense CIF will be less than that of those in the less dense CIF*, by equivalence of the two situations.

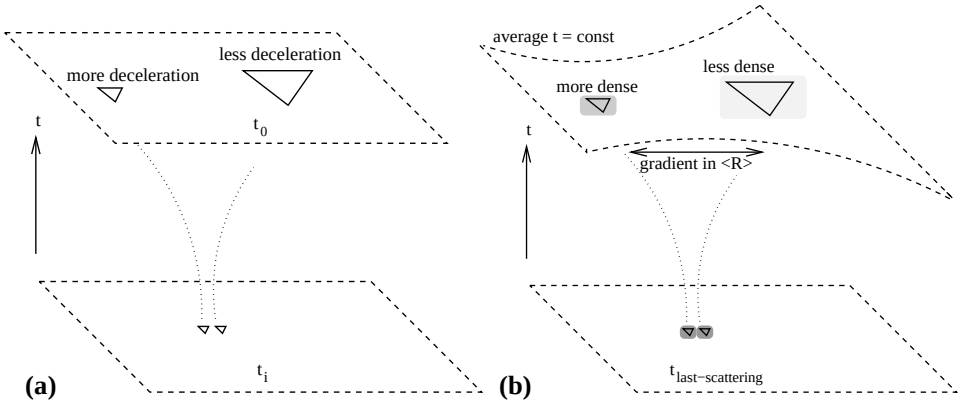


FIGURE 1. Two equivalent situations: **(a)** in Minkowski space observers in separate semi–tethered lattices, initially expanding at the same rate, apply brakes homogeneously and isotropically within their respective regions but at different rates; **(b)** in the universe which is close to homogeneous and isotropic at last-scattering comoving observers in separated regions initially move away from each other isotropically, but experience different locally homogeneous isotropic decelerations as local density contrasts grow. In both cases there is a relative deceleration of the observer congruences and those in the region which has decelerated more will age less.

The fact that a global Minkowski observer does not exist in the second case does not invalidate the argument. The global Minkowski time is just a coordinate label. In the cosmological case the only restriction is that *the expansion of both average congruences must remain homogeneous and isotropic in local regions of different average density* in the global average $t =$const slice. Provided we can patch the regional frames together suitably, then if regions in such a slice *are still expanding* and have a significant density contrast we can expect a significant clock rate variance.

This equivalence directly establishes the idea of a *gravitational energy cost for a spatial curvature gradient*, since the existence of expanding regions of different density within an average $t =$const slice implies a gradient in the average Ricci scalar curvature, $\langle \mathscr{R} \rangle$, on one hand, while the fact that the local proper time varies on account of the relative deceleration implies a gradient in gravitational energy on the other.

In the actual universe, the question is: can the effect described above be significant enough to give a significant variation in the clocks of ideal isotropic observers (those who see an isotropic mean CMB) in regions of different density, who experience a relative deceleration of their regional volume expansions? Since we are dealing with weak fields the relative deceleration of the background is small. Nonetheless even if the relative deceleration is typically of order $10^{-10}\mathrm{ms}^{-2}$, cumulatively over the age of the universe it leads to significant clock rate variances [3], of the order of 38%. Such a large effect is counterintuitive, as we are used to only considering time dilations due to relative accelerations within the static potentials of isolated systems. Essentially, we are dealing with a different physical effect concerning the relative synchronization of clocks in the absence of global Killing vectors. A small instantaneous relative deceleration can lead to cumulatively large differences, given one has the lifetime of the universe to play with. As a consequence the age of the universe itself becomes position–dependent. Since we and all the objects we observe are necessarily in regions of greater than critical density, on

account of structure formation we have a mass–biased view of the universe and cannot directly observe such variations.

In the standard ADM formalism one assumes the existence of a global rest frame comoving with the dust, and one makes a $(3+1)$–split of the Einstein equations from the point of view of fundamental observers who may be either comoving or tilted with respect to the dust. If the only symmetries that are allowed are diffeomorphisms of the global metric on one hand, and local Lorentz transformations corresponding to rotations and boosts on the other, then realistically there is no room within such an ADM formalism for clock rate variations of the order of magnitude dealt with in the timescape scenario [18]. However, such criticism overlooks the very real possibility the rest frame of dust is not globally defined, and furthermore it overlooks the crucial idea of regional averages introduced by the CEP. In proposing to separate the collective degree of freedom of the quasi-local regional volume expansion from other gravitational degrees of freedom, I am suggesting that we must consider average regional symmetries as a completely new ingredient in addition to the global diffeomorphisms and local Lorentz transformations with which we are familiar. This is a potential route to dealing with the unsolved problems of coarse-graining in general relativity.

At the epoch of last scattering dust may certainly be assumed to be atomic. However, once structures form geodesics cross and at the present epoch dust must be coarse-grained on at least the scale of galaxies in cosmological modelling. Thus the issue of the coarse-graining of dust is not merely a matter of choice, but of physical necessity if one is to consistently think about the interpretation of the Buchert formalism[3]. Of course, a detailed mathematical framework[4] for this still remains to be given in the timescape scenario. However, it is my view that mathematics is best guided by physical intuition rather than the reverse, and consequently my work to date has proceeded from making a phenomenological ansatz consistent with the CEP, to see whether the idea stands a chance of working.

THE TIMESCAPE MODEL

I proceed from an ansatz that the variance in gravitational energy is correlated with the average spatial curvature in such a way as to implicitly solve the Sandage–de Vaucouleurs paradox that a statistically quiet, broadly isotropic, Hubble flow is observed deep below the scale of statistical homogeneity. In particular, galaxy peculiar velocities have a small magnitude with respect to a local regional volume expansion. Expanding regions of different densities are patched together so that the regionally measured expansion remains uniform. Such regional expansion refers to the variation of the regional

---

[3] One can apply the Buchert formalism in a different manner – for example, to the problem of prescribed dust in exact solutions such as the LTB model [19]–[21] on globally well-defined spacelike hypersurfaces, where one specifically avoids solutions which develop vorticity or singularities. However, it is my view that to deal with the actual inhomogeneities of the observed universe then the average evolution of the Einstein equations should be regarded as a statistical description, and I approach the Buchert formalism in this sense.

[4] For one approach to coarse-graining, as opposed to averaging, see ref. [22].

proper length, $\ell_r = \mathscr{V}^{1/3}$, with respect to proper time of isotropic observers. Although voids open up faster, so that their proper volume increases more quickly, on account of gravitational energy gradients the local clocks will also tick faster in a compensating manner.

In order to deal with dust evolution from the surface of last scattering up to the present epoch, I assume that dust can be coarse–grained at the $100h^{-1}$Mpc scale of statistical homogeneity over which mass flows can be neglected. The manner in which I interpret the Buchert formalism is therefore different to that adopted by Buchert [12], who does not define the scale of coarse-graining of the dust explicitly. Details of the fitting of local observables to average quantities for solutions to the Buchert equations[5] are described in detail in refs. [2, 6]. Negatively curved voids, and spatially flat expanding wall regions within which galaxy clusters are located, are combined in a Buchert average

$$f_{\rm v}(t) + f_{\rm w}(t) = 1, \tag{2}$$

where $f_{\rm w}(t) = f_{\rm wi}a_{\rm w}^3/\bar{a}^3$ is the *wall volume fraction* and $f_{\rm v}(t) = f_{\rm vi}a_{\rm v}^3/\bar{a}^3$ is the *void volume fraction*, $\mathscr{V} = \mathscr{V}_{\rm i}\bar{a}^3$ being the present horizon volume, and $f_{\rm wi}$, $f_{\rm vi}$ and $\mathscr{V}_{\rm i}$ initial values at last scattering. The time parameter, t, is the volume–average time parameter of the Buchert formalism, but does not coincide with that of local measurements in galaxies. In trying to fit a FLRW solution to the universe we attempt to match our local spatially flat wall geometry

$$\mathrm{d}s^2_{\rm fl} = -\mathrm{d}\tau^2 + a_{\rm w}^2(\tau)\left[\mathrm{d}\eta_{\rm w}^2 + \eta_{\rm w}^2\mathrm{d}\Omega^2\right]. \tag{3}$$

to the whole universe, when in reality the calibration of rulers and clocks of ideal isotropic observers vary with gradients in spatial curvature and gravitational energy. By conformally matching radial null geodesics with those of the Buchert average solutions, the geometry (3) may be extended to cosmological scales as the dressed geometry

$$\mathrm{d}s^2 = -\mathrm{d}\tau^2 + a^2(\tau)\left[\mathrm{d}\bar{\eta}^2 + r_{\rm w}^2(\bar{\eta},\tau)\,\mathrm{d}\Omega^2\right] \tag{4}$$

where $a = \bar{\gamma}^{-1}\bar{a}$, $\bar{\gamma} = \frac{\mathrm{d}t}{\mathrm{d}\tau}$ is the relative lapse function[6] between wall clocks and volume–average ones, $\mathrm{d}\bar{\eta} = \mathrm{d}t/\bar{a} = \mathrm{d}\tau/a$, and $r_{\rm w} = \bar{\gamma}(1-f_{\rm v})^{1/3}\,{f_{\rm wi}}^{-1/3}\eta_{\rm w}(\bar{\eta},\tau)$, where $\eta_{\rm w}$ is given by integrating $\mathrm{d}\eta_{\rm w} = {f_{\rm wi}}^{1/3}\mathrm{d}\bar{\eta}/[\bar{\gamma}(1-f_{\rm v})^{1/3}]$ along null geodesics.

In addition to the bare cosmological parameters which describe the Buchert equations, one obtains dressed parameters relative to the geometry (4). For example, the dressed matter density parameter is $\Omega_M = \bar{\gamma}^3\bar{\Omega}_M$, where $\bar{\Omega}_M = 8\pi G\bar{\rho}_{M0}\bar{a}_0^3/(3\bar{H}^2\bar{a}^3)$ is the bare

---

[5] The model of Wiegand and Buchert [23], briefly described by Buchert in the present volume [24], has similarities to the present model but also differs from it in certain key aspects. In particular, (i) the observational interpretation of the Buchert averages is different; (ii) the walls and voids are taken to have internal backreaction in the case or refs. [23, 24] but not here; and (iii) the interpretation of the initial wall and void fractions at last scattering is different. In respect of the last point, since walls and voids do not exist at the surface of last scattering, I take the view that the vast bulk of the present horizon volume that averages to critical density gives $f_{\rm wi} \simeq 1$, while $f_{\rm vi} = 1 - f_{\rm wi}$ is the small positive fraction of the present horizon volume that consists of uncompensated underdense regions at last scattering surface.

[6] This is a phenomenological function rather than the lapse function prescribed by the ADM formalism.

matter density parameter. The dressed parameters take numerical values close to the ones inferred in standard FLRW models.

Apparent acceleration and Hubble flow variance

The gradient in gravitational energy and cumulative differences of clock rates between wall observers and volume average ones has important physical consequences. Using the exact solution obtained in ref. [6], one finds that a volume average observer would infer an effective deceleration parameter $\bar{q} = -\ddot{\bar{a}}/(\bar{H}^2\bar{a}) = 2\,(1-f_{\rm v})^2\,/(2+f_{\rm v})^2$, which is always positive since there is no global acceleration. However, a wall observer infers a dressed deceleration parameter

$$q = \frac{-1}{H^2 a}\frac{{\rm d}^2 a}{{\rm d}\tau^2} = \frac{-\,(1-f_{\rm v})\,(8{f_{\rm v}}^3+39{f_{\rm v}}^2-12f_{\rm v}-8)}{\left(4+f_{\rm v}+4{f_{\rm v}}^2\right)^2}\,, \tag{5}$$

where the dressed Hubble parameter is given by

$$H = a^{-1}\,\tfrac{\rm d}{{\rm d}\tau}\,a = \bar{\gamma}\bar{H} - \dot{\bar{\gamma}} = \bar{\gamma}\bar{H} - \bar{\gamma}^{-1}\,\tfrac{\rm d}{{\rm d}\tau}\,\bar{\gamma}. \tag{6}$$

At early times when $f_{\rm v}\to 0$ the dressed and bare deceleration parameter both take the Einstein–de Sitter value $q\simeq\bar{q}\simeq\frac{1}{2}$. However, unlike the bare parameter which monotonically decreases to zero, the dressed parameter becomes negative when $f_{\rm v}\simeq 0.59$ and $\bar{q}\to 0^-$ at late times. For the best-fit parameters[7] the apparent acceleration begins at a redshift $z\simeq 0.9$.

Cosmic acceleration is thus revealed as an apparent effect which arises due to the cumulative clock rate variance of wall observers relative to volume–average observers. It becomes significant only when the voids begin to dominate the universe by volume. Since the epoch of onset of apparent acceleration is directly related to the void fraction, $f_{\rm v}$, this solves one cosmic coincidence problem.

In addition to apparent cosmic acceleration, a second important apparent effect will arise if one considers scales below that of statistical homogeneity. By any one set of clocks it will appear that voids expand faster than wall regions. Thus a wall observer will see galaxies on the far side of a dominant void of diameter $30h^{-1}$ Mpc recede at a rate greater than the dressed global average H_0, while galaxies within an ideal wall will recede at a rate less than H_0. Since the bare Hubble parameter $\bar{H}$ provides a measure of the uniform quasi-local flow, it must also be the "local value" within an ideal wall at any epoch; i.e., eq. (6) gives a measure of the variance in the apparent Hubble flow. The best-fit parameters [25] give a dressed Hubble constant $H_0 = 61.7^{+1.2}_{-1.1}\,{\rm km\ sec^{-1}\,Mpc^{-1}}$, and a bare Hubble constant $\bar{H}_0 = 48.2^{+2.0}_{-2.4}\,{\rm km\ sec^{-1}\,Mpc^{-1}}$. The present epoch variance is 17–22%.

---

[7] Here I will simply adopt the parameters found in ref. [25] from a fit to the Riess07 gold dataset [26]. A more recent analysis [27] shows that the best-fit parameters are sensitive to the method of supernova data reduction, and unknown systematic issues remain to be resolved. The parameters determined from the Riess07 dataset are in the mid–range of those determined by MLCS methods from larger datasets [27].

Since voids dominate the universe by volume at the present epoch, any observer in a galaxy in a typical wall region will measure locally higher values of the Hubble constant, with peak values of order 72 km sec^{-1} Mpc^{-1} at the $30h^{-1}$ Mpc scale of the dominant voids. Over larger distances, as the line of sight intersects more walls as well as voids, a radial spherically symmetric average will give an average Hubble constant whose value decreases from the maximum at the $30h^{-1}$ Mpc scale to the dressed global average value, as the scale of homogeneity is approached at roughly the baryon acoustic oscillation (BAO) scale of $110h^{-1}$Mpc. This predicted effect could account for the Hubble bubble [28] and more detailed studies of the scale dependence of the local Hubble flow [29].

In fact, the variance of the local Hubble flow below the scale of homogeneity should correlate strongly to observed structures in a manner which has no equivalent prediction in FLRW models.

There is already evidence from the study of large–scale bulk flows that apparent peculiar velocities determined in the FLRW framework have a magnitude in excess of the expectations of the standard ΛCDM model [30, 31]. In the present framework, rather than having a uniform expansion (with respect to one set of clocks), with respect to which peculiar velocities are defined, we have variations in the expansion rate in regions of different density which are expanding but decelerating at different rates. Nonetheless, given that our location is on the edge of a dominant void and a wall [32] the equivalent maximum peculiar velocity can be estimated as

$$v_{\rm pec} = (\tfrac{3}{2}\bar{H}_0 - H_0)\frac{30}{h}\,\text{Mpc} = 510^{+210}_{-260}\,\text{km/s} \tag{7}$$

assuming a diameter of $30\,h^{-1}$ Mpc for the local dominant void. This rough estimate is of a magnitude consistent with observation.

FUTURE OBSERVATIONAL TESTS

There are two types of potential cosmological tests that can be developed; those relating to scales below that of statistical homogeneity as discussed above, and those that relate to averages on our past light cone on scales much greater than the scale of statistical homogeneity. The second class of tests includes equivalents to all the standard cosmological tests of the standard FLRW model with Newtonian perturbations. This second class of tests can be further divided into tests which just deal with the bulk cosmological averages (luminosity and angular diameter distances etc), and those that deal with the variance from the growth of structures (late epoch integrated Sachs–Wolfe effect, cosmic shear, redshift space distortions etc). Here I will concentrate solely on the simplest tests which are directly related to luminosity and angular diameter distance measures.

In the timescape cosmology we have an effective dressed luminosity distance

$$d_L = a_0(1+z)r_{\rm w}, \tag{8}$$

where $a_0 = \bar{\gamma}_0^{-1}\bar{a}_0$, and

$$r_{\rm w} = \bar{\gamma}(1-f_{\rm v})^{1/3}\int_t^{t_0}\frac{{\rm d}t'}{\bar{\gamma}(t')(1-f_{\rm v}(t'))^{1/3}\bar{a}(t')}\,. \tag{9}$$

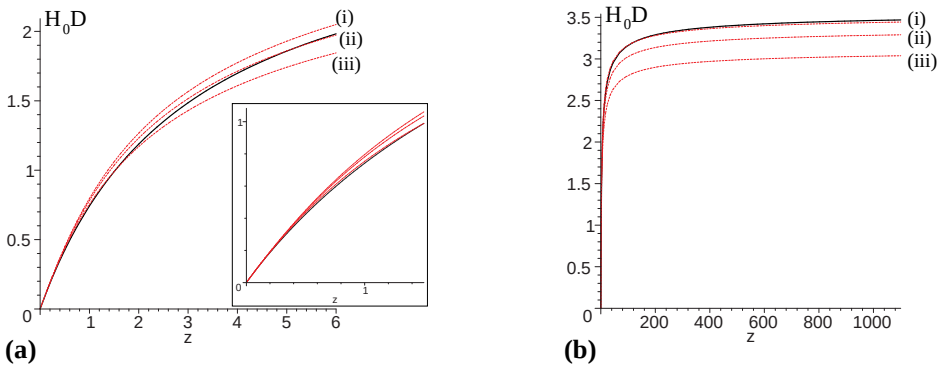


FIGURE 2. The effective comoving distance $H_0D(z)$ is plotted for the best-fit timescape (TS) model, with $f_{\rm v0} = 0.762$, (solid line); and for various spatially flat ΛCDM models (dashed lines). The parameters for the dashed lines are (i) $\Omega_{M0} = 0.249$ (best-fit to WMAP5 only [33]); (ii) $\Omega_{M0} = 0.279$ (joint best-fit to SneIa, BAO and WMAP5); (iii) $\Omega_{M0} = 0.34$ (best-fit to Riess07 SneIa only [26]). Panel **(a)** shows the redshift range $z < 6$, with an inset for $z < 1.5$, which is the range tested by current SneIa data. Panel **(b)** shows the range $z < 1100$ up to the surface of last scattering, tested by WMAP.

We can also define an *effective angular diameter distance*, d_A, and an *effective comoving distance*, D, to a redshift z in the standard fashion

$$d_A = \frac{D}{1+z} = \frac{d_L}{(1+z)^2}\,. \tag{10}$$

A direct method of comparing the distance measures with those of homogeneous models with dark energy, is to observe that for a standard spatially flat cosmology with dark energy obeying an equation of state $P_D = w(z)\rho_D$, the quantity

$$H_0D = \int_0^z \frac{{\rm d}z'}{\sqrt{\Omega_{M0}(1+z')^3 + \Omega_{D0}\exp\left[3\int_0^{z'}\frac{(1+w(z''))dz''}{1+z''}\right]}}\,, \tag{11}$$

does not depend on the value of the Hubble constant, H_0, but only directly on $\Omega_{M0} = 1 - \Omega_{D0}$. Since the best-fit values of H_0 are potentially different for the different scenarios, a comparison of H_0D curves as a function of redshift for the timescape model versus the ΛCDM model gives a good indication of where the largest differences can be expected, independently of the value of H_0. Such a comparison is made in Fig. 2.

We see that as redshift increases the timescape model interpolates between ΛCDM models with different values of Ω_{M0}. For redshifts $z \lesssim 1.5$ $D_{\rm TS}$ is very close to $D_{\Lambda{\rm CDM}}$ for the parameter values $(\Omega_{M0}, \Omega_{\Lambda 0}) = (0.34, 0.66)$ (model (iii)) which best-fit the Riess07 supernovae (SneIa) data [26] only, by our own analysis. For very large redshifts that approach the surface of last scattering, $z \lesssim 1100$, on the other hand, $D_{\rm TS}$ very closely matches $D_{\Lambda{\rm CDM}}$ for the parameter values $(\Omega_{M0}, \Omega_{\Lambda 0}) = (0.249, 0.751)$ (model (i)) which best-fit WMAP5 only [33]. Over redshifts $2 \lesssim z \lesssim 10$, at which scales independent tests are conceivable, $D_{\rm TS}$ makes a transition over corresponding curves of $D_{\Lambda{\rm CDM}}$ with intermediate values of $(\Omega_{M0}, \Omega_{\Lambda 0})$. The $D_{\Lambda{\rm CDM}}$ curve for joint best-fit parameters to SneIa, BAO measurements and WMAP5 [33], $(\Omega_{M0}, \Omega_{\Lambda 0}) = (0.279, 0.721)$ is best–matched over the range $5 \lesssim z \lesssim 6$, for example.

The difference of $D_{\rm TS}$ from any single $D_{\Lambda{\rm CDM}}$ curve is perhaps most pronounced in the range $2 \lesssim z \lesssim 6$, which may be an optimal regime to probe in future experiments. Gamma–ray bursters (GRBs) now probe distances to redshifts $z \lesssim 8.3$, and could be very useful if their properties could be understood to the extent that they might be reliably used as standard candles. A considerable amount work of work has already been done on Hubble diagrams for GRBs. (See, e.g., [34].) Much more work is needed to nail down systematic uncertainties, but GRBs may eventually provide a definitive test in future. An analysis of the timescape model Hubble diagram using 69 GRBs has just been performed by Schaefer [35], who finds that the timescape model fits the data better than the concordance ΛCDM model, but not yet by a huge margin[8]. As more data is accumulated, it should become possible to distinguish the models if the issues with the standardization of GRBs can be ironed out.

The effective "equation of state"

The shape of the H_0D curves depicted in Fig. 2 represents the observable quantity one is actually measuring in tests some researchers loosely refer to as "measuring the equation of state". For spatially flat dark energy models, with H_0D given by (11), one finds that the function $w(z)$ appearing in the fluid equation of state $P_D = w(z)\rho_D$ is related to the first and second derivatives of (11) by

$$w(z) = \frac{\frac{2}{3}(1+z)D'^{-1}D'' + 1}{\Omega_{M0}(1+z)^3 H_0^2 D'^2 - 1} \tag{12}$$

where prime denotes a derivative with respect to z. Such a relation can be applied to observed distance measurements, regardless of whether the underlying cosmology has dark energy or not. Since it involves first and second derivatives of the observed quantities, it is actually much more difficult to determine observationally than directly fitting $H_0D(z)$.

The equivalent of the "equation of state", $w(z)$, for the timescape model is plotted in Fig. 3. The fact that $w(z)$ is undefined at a particular redshift and changes sign through $\pm\infty$ simply reflects the fact that in (12) we are dividing by a quantity which goes to zero for the timescape model, even though the underlying curve of Fig. 2 is smooth. Since one is not dealing with a dark energy fluid in the present case, $w(z)$ simply has no physical meaning. Nonetheless, phenomenologically the results do agree with the usual inferences about $w(z)$ for fits of standard dark energy cosmologies to SneIa data. For the canonical model of Fig. 3(a) one finds that the average value of $w(z) \simeq -1$ on the range $z \lesssim 0.7$, while the average value of $w(z) < -1$ if the range of redshifts is extended to higher values. The $w = -1$ "phantom divide" is crossed at $z \simeq 0.46$ for $f_{\rm v0} \simeq 0.76$. One recent study [37] finds mild 95% evidence for an equation of state that crosses the phantom divide from $w > -1$ to $w < -1$ in the range $0.25 < z < 0.75$ in accord with the

[8] By contrast the conformal gravity model of Mannheim [36] produced a worse fit, while the Chaplygin gas fit best only in the limit that its parameters reduce to those of the ΛCDM model [35].

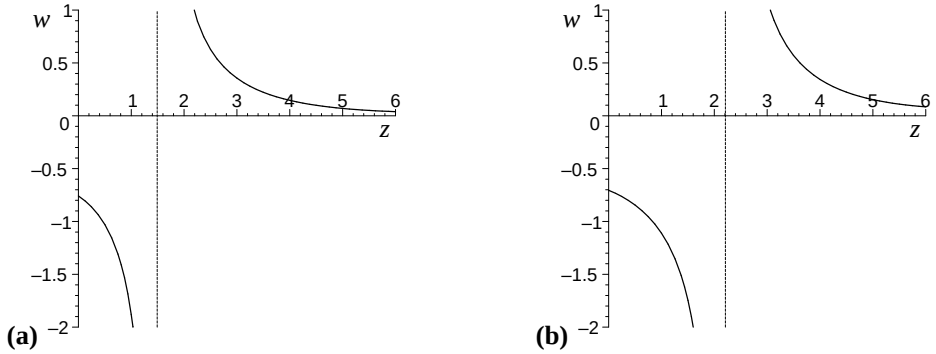


FIGURE 3. The artificial equivalent of an equation of state constructed using the effective comoving distance (12), plotted for the timescape tracker solution with best-fit value $f_{\rm v0} = 0.762$, and two different values of Ω_{M0}: **(a)** the canonical dressed value $\Omega_{M0} = \frac{1}{2}(1 - f_{\rm v0})(2 + f_{\rm v0}) = 0.33$; **(b)** $\Omega_{M0} = 0.279$.

timescape expectation. By contrast, another study [38] at redshifts $z < 1$ draws different conclusions about dynamical dark energy, but for the given uncertainties in $w(z)$ the data is consistent with Fig. 2(a) as well as with a cosmological constant [5].

The fact that $w(z)$ is a different sign to the dark energy case for $z > 2$ is another way of viewing our statement above that the redshift range $2 \lesssim z \lesssim 6$ may be optimal for discriminating model differences.

The $H(z)$ measure

Further observational diagnostics can be devised if the expansion rate $H(z)$ can be observationally determined as a function of redshift. Recently such a determination of $H(z)$ at $z = 0.24$ and $z = 0.43$ has been made using redshift space distortions of the BAO scale in the ΛCDM model [39]. This technique is of course model dependent, and the Kaiser effect would have to be re-examined in the timescape model before a direct comparison of observational results could be made. A model–independent measure of $H(z)$, the redshift time drift test, is discussed below.

In Fig. 4 we compare $H(z)/H_0$ for the timescape model to spatially flat ΛCDM models with the same parameters chosen in Fig. 2. The most notable feature is that the slope of $H(z)/H_0$ is less than in the ΛCDM case, as is to be expected for a model whose (dressed) deceleration parameter varies more slowly than for ΛCDM.

The $Om(z)$ measure

Recently a number of authors [40, 41, 42] have discussed various roughly equivalent diagnostics of dark energy. For example, Sahni, Shafieloo and Starobinsky [41], have proposed a diagnostic function

$$Om(z) = \left[\frac{H^2(z)}{H_0^2} - 1\right] \left[(1+z)^3 - 1\right]^{-1} , \qquad (13)$$

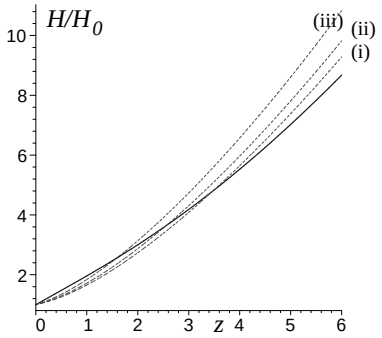


FIGURE 4. The function $H_0^{-1}H(z)$ for the timescape model with $f_{\rm v0} = 0.762$ (solid line) is compared to $H_0^{-1}H(z)$ for three spatially flat ΛCDM models with the same values of $(\Omega_{M0}, \Omega_{\Lambda 0})$ as in Fig. 2 (dashed lines).

on account of the fact that it is equal to the constant present epoch matter density parameter, Ω_{M0}, at all redshifts for a spatially flat FLRW model with pressureless dust and a cosmological constant. However, it is not constant if the cosmological constant is replaced by other forms of dark energy. For general FLRW models, $H(z) = [D'(z)]^{-1}\sqrt{1+\Omega_{k0}H_0^2D^2(z)}$, which only involves a single derivatives of $D(z)$. Thus the diagnostic (13) is easier to reconstruct observationally than the equation of state parameter, $w(z)$.

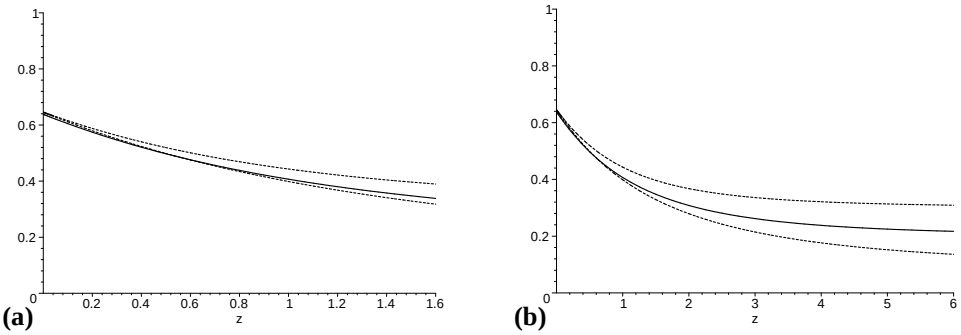


FIGURE 5. The dark energy diagnostic $Om(z)$ of Sahni, Shafieloo and Starobinsky [41] plotted for the timescape tracker solution with best-fit value $f_{\rm v0} = 0.762$ (solid line), and 1σ limits (dashed lines) from ref. [25]: **(a)** for the redshift range $0 < z < 1.6$ as shown in ref. [43]; **(b)** for the redshift range $0 < z < 6$.

The quantity $Om(z)$ is readily calculated for the timescape model, and the result is displayed in Fig. 5. What is striking about Fig. 5, as compared to the curves for quintessence and phantom dark energy models as plotted in ref. [41], is that the initial value

$$Om(0) = \tfrac{2}{3}\,H'\big|_0 = \frac{2(8f_{\rm v0}^3 - 3f_{\rm v0}^2 + 4)(2+f_{\rm v0})}{(4f_{\rm v0}^2 + f_{\rm v0} + 4)^2} \qquad (14)$$

is substantially larger than in the spatially flat dark energy models. Furthermore, for the timescape model $Om(z)$ does not asymptote to the dressed density parameter Ω_{M0} in any redshift range. For quintessence models $Om(z) > \Omega_{M0}$, while for phantom models $Om(z) < \Omega_{M0}$, and in both cases $Om(z) \to \Omega_{M0}$ as $z \to \infty$. In the timescape model,

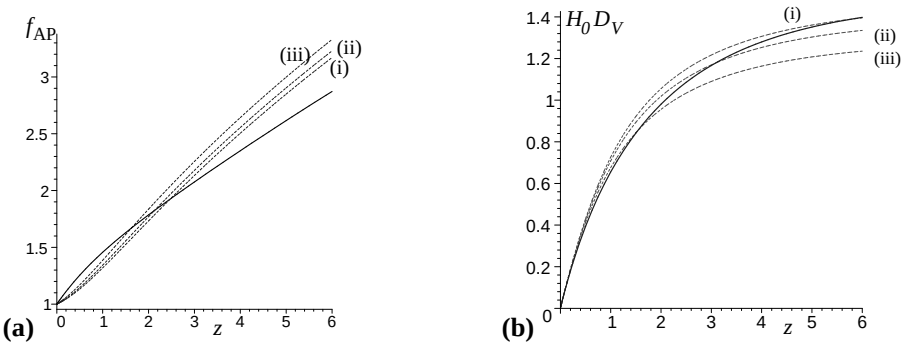


FIGURE 6. **(a)** The Alcock–Paczyński test function $f_{\rm AP} = HD/z$; and **(b)** the BAO radial test function $H_0 D_V = H_0 D f_{\rm AP}^{-1/3}$. In each case the timescape model with $f_{\rm v0} = 0.762$ (solid line) is compared to three spatially flat ΛCDM models with the same values of $(\Omega_{M0}, \Omega_{\Lambda 0})$ as in Fig. 2 (dashed lines).

$Om(z) > \Omega_{M0} \simeq 0.33$ for $z \lesssim 1.7$, while $Om(z) < \Omega_{M0}$ for $z \gtrsim 1.7$. It thus behaves more like a quintessence model for low z, in accordance with Fig. 3. However, the steeper slope and the different large z behaviour mean the diagnostic is generally very different to that of typical dark energy models. For large z, $\Omega_{M0} < Om(\infty) < \Omega_{M0}$, if $f_{\rm v0} > 0.25$.

Interestingly enough, a recent analysis of SneIa, BAO and CMB data [43] for dark energy models with two different empirical fitting functions for $w(z)$ gives an intercept $Om(0)$ which is larger than expected for typical quintessence or phantom energy models, and in the better fit of the two models the intercept (see Fig. 3 of ref. [43])] is close to the value expected for the timescape model, which is tightly constrained to the range $0.638 < Om(0) < 0.646$ if $f_{\rm v0} = 0.76^{+0.12}_{-0.09}$.

The Alcock–Paczyński test and baryon acoustic oscillations

Some time ago Alcock and Paczyński devised a test [44] which relies on comparing the radial and transverse proper length scales of spherical standard volumes comoving with the Hubble flow. This test, which determines the function

$$f_{\rm AP} = \frac{1}{z} \left| \frac{\delta\theta}{\delta z} \right| = \frac{HD}{z}, \tag{15}$$

was originally conceived to distinguish FLRW models with a cosmological constant from those without a Λ term. The test is free from many evolutionary effects, but relies on one being able to remove systematic distortions due to peculiar velocities.

Current detections of the BAO scale in galaxy clustering statistics [45, 46] can in fact be viewed as a variant of the Alcock–Paczyński test, as they make use of both the transverse and radial dilations of the fiducial comoving BAO scale to present a measure

$$D_V = \left[\frac{zD^2}{H(z)} \right]^{1/3} = D f_{\rm AP}^{-1/3}. \tag{16}$$

In Fig. 6 the Alcock–Paczyński test function (15) and BAO scale measure (16) of the timescape model are compared to those of the spatially flat ΛCDM model with different values of ($\Omega_{\Lambda 0}$,$\Omega_{\Lambda 0}$). Over the range of redshifts $z < 1$ studied currently with galaxy clustering statistics, the $f_{\rm AP}$ curve distinguishes the timescape model from the ΛCDM models much more strongly than the D_V test function. In particular, the timescape $f_{\rm AP}$ has a distinctly different shape to that of the ΛCDM model, being convex. The primary reason for use of the integral measure (16) has been a lack of data. Future measurements with enough data to separate the radial and angular BAO scales are a potentially powerful way of distinguishing the timescape model from ΛCDM.

Recently Gaztañaga, Cabré and Hui [39] have made the first efforts to separate the radial and angular BAO scales in different redshift slices. Although they have not yet published separate values for the radial and angular scales, their results are interesting when compared to the expectations of the timescape model. Their study yields best-fit values of the present total matter and baryonic matter density parameters, Ω_{M0} and Ω_{B0}, which are in tension with WMAP5 parameters fit to the ΛCDM model. In particular, the ratio of nonbaryonic cold dark matter to baryonic matter has a best-fit value $\Omega_{C0}/\Omega_{B0} = (\Omega_{M0} - \Omega_{B0})/\Omega_{B0}$ of 3.7 in the $0.15 < z < 0.3$ sample, 2.6 in the $0.4 < z < 0.47$ sample, and 3.6 in the whole sample, as compared to the expected value of 6.1 from WMAP5. The analysis of the 3–point correlation function yields similar conclusions, with a best fit [47] $\Omega_{M0} = 0.28 \pm 0.05$, $\Omega_{B0} = 0.079 \pm 0.025$. By comparison, the parameter fit to the timescape model of ref. [25] yields dressed parameters $\Omega_{M0} = 0.33^{+0.11}_{-0.16}$, $\Omega_{B0} = 0.080^{+0.021}_{-0.013}$, and a ratio $\Omega_{C0}/\Omega_{B0} = 3.1^{+2.5}_{-2.4}$. Since homogeneous dark energy models are not generally expected to give rise to a renormalization of the ratio of nonbaryonic to baryonic matter, this is encouraging for the timescape model.

Test of (in)homogeneity

Recently Clarkson, Bassett and Lu [48] have constructed what they call a "test of the Copernican principle" based on the observation that for homogeneous, isotropic models which obey the Friedmann equation, the present epoch curvature parameter, a constant, may be written as

$$\Omega_{k0} = \frac{[H(z)D'(z)]^2 - 1}{[H_0 D(z)]^2} \tag{17}$$

for all z, irrespective of the dark energy model or any other model parameters. Consequently, taking a further derivative, the quantity

$$\mathscr{C}(z) \equiv 1 + H^2(DD'' - D'^2) + HH'DD' \tag{18}$$

must be zero for all redshifts for any FLRW geometry.

A deviation of $\mathscr{C}(z)$ from zero, or of (17) from a constant value, would therefore mean that the assumption of homogeneity is violated. Although this only constitutes a test of the assumption of the Friedmann equation, i.e., of the Cosmological Principle rather than the broader Copernican Principle adopted in ref. [2], the average inhomogeneity will give a clear and distinct prediction of a nonzero $\mathscr{C}(z)$ for the timescape model.

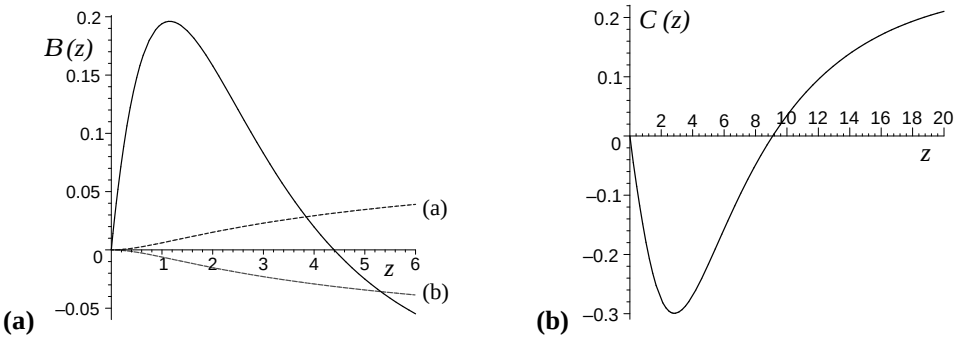


FIGURE 7. Left panel: The (in)homogeneity test function $\mathscr{B}(z) = [HD']^2 - 1$ is plotted for the timescape tracker solution with best-fit value $f_{\rm v0} = 0.762$ (solid line), and compared to the equivalent curves $\mathscr{B} = \Omega_{k0}(H_0 D)^2$ for two different ΛCDM models with small curvature: **(a)** $\Omega_{M0} = 0.28$, $\Omega_{\Lambda 0} = 0.71$, $\Omega_{k0} = 0.01$; **(b)** $\Omega_{M0} = 0.28$, $\Omega_{\Lambda 0} = 0.73$, $\Omega_{k0} = -0.01$.
Right panel: The (in)homogeneity test function $\mathscr{C}(z)$ is plotted for the $f_{\rm v0} = 0.762$ tracker solution.

The functions (17) and (18) are computed in ref. [5]. Observationally it is more feasible to fit (17) which involves one derivative less of redshift. In Fig. 7 we exhibit both $\mathscr{C}(z)$, and also the function $\mathscr{B}(z) = [HD']^2 - 1$ from the numerator of (17) for the timescape model, as compared to two ΛCDM models with a small amount of spatial curvature. A spatially flat FLRW model would have $\mathscr{B}(z) \equiv 0$. In other FLRW cases $\mathscr{B}(z)$ is always a monotonic function whose sign is determined by that of Ω_{k0}. An open $\Lambda = 0$ universe with the same Ω_{M0} would have a monotonic function $\mathscr{B}(z)$ very much greater than that of the timescape model.

Time drift of cosmological redshifts

For the purpose of the $Om(z)$ and (in)homogeneity tests considered in the last section, $H(z)$ must be observationally determined, and this is difficult to achieve in a model-independent way. There is one way of achieving this, however, namely by measuring the time variation of the redshifts of different sources over a sufficiently long time interval [49], as has been discussed recently by Uzan, Clarkson and Ellis [50]. Although the measurement is extremely challenging, it may be feasible over a 20 year period by precision measurements of the Lyman-α forest in the redshift range $2 < z < 5$ with the next generation of Extremely Large Telescopes [51].

In ref. [5] an analytic expression for $H_0^{-1}\frac{{\rm d}z}{{\rm d}\tau}$ is determined, the derivative being with respect to wall time for observers in galaxies. The resulting function is displayed in Fig. 8 for the best-fit timescape model with $f_{\rm v0} = 0.762$, where it is compared to the equivalent function for three different spatially flat ΛCDM models. What is notable is that the curve for the timescape model is considerably flatter than those of the ΛCDM models. This may be understood to arise from the fact that the magnitude of the apparent acceleration is considerably smaller in the timescape model, as compared to the magnitude of the acceleration in ΛCDM models. For models in which there is no apparent acceleration whatsoever, one finds that $H_0^{-1}\frac{{\rm d}z}{{\rm d}\tau}$ is always negative. If there is cosmic acceleration at

late epochs, real or apparent, then $H_0^{-1}\frac{\mathrm{d}z}{\mathrm{d}\tau}$ will become positive at low redshifts, though at a somewhat larger redshift than that at which acceleration is deemed to have begun.

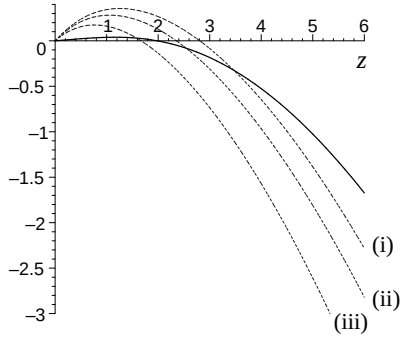


FIGURE 8. The function $H_0^{-1}\frac{\mathrm{d}z}{\mathrm{d}\tau}$ for the timescape model with $f_{\mathrm{v}0} = 0.762$ (solid line) is compared to $H_0^{-1}\frac{\mathrm{d}z}{\mathrm{d}\tau}$ for three spatially flat ΛCDM models with the same values of $(\Omega_{M0}, \Omega_{\Lambda 0})$ as in Fig. 2 (dashed lines).

Fig. 8 demonstrates that a very clear signal of differences in the redshift time drift between the timescape model and ΛCDM models might be determined at low redshifts when $H_0^{-1}\frac{\mathrm{d}z}{\mathrm{d}\tau}$ should be positive. In particular, the magnitude of $H_0^{-1}\frac{\mathrm{d}z}{\mathrm{d}\tau}$ is considerably smaller for the timescape model as compared to ΛCDM models. Observationally, however, it is expected that measurements will be best determined for sources in the Lyman α forest in the range, $2 < z < 5$. At such redshifts the magnitude of the drift is somewhat more pronounced in the case of the ΛCDM models. For a source at $z = 4$, over a period of $\delta\tau = 10$ years we would have $\delta z = -3.3 \times 10^{-10}$ for the timescape model with $f_{\mathrm{v}0} = 0.762$ and $H_0 = 61.7\,\mathrm{km\ sec^{-1}\,Mpc^{-1}}$. By comparison, for a spatially flat ΛCDM model with $H_0 = 70.5\,\mathrm{km\ sec^{-1}\,Mpc^{-1}}$ a source at $z = 4$ would over ten years give $\delta z = -4.7 \times 10^{-10}$ for $(\Omega_{M0}, \Omega_{\Lambda 0}) = (0.249, 0.751)$, and $\delta z = -7.0 \times 10^{-10}$ for $(\Omega_{M0}, \Omega_{\Lambda 0}) = (0.279, 0.721)$.

DARK MATTER AND THE TIMESCAPE SCENARIO

Since much of this conference has been about alternatives to standard nonbaryonic dark matter, I will briefly comment on the issue of dark matter vis–à–vis the timescape scenario. The timescape model only addresses large scale cosmological averages, and does not make specific predictions about dark matter typically inferred from direct observations of bound systems, such as rotation curves of galaxies, gravitational lensing or motions of galaxies within clusters. However, as has been discussed in ref. [2], in a re-examination of the post-Newtonian approximation within the averaging problem in cosmology, departures from the naïve Newtonian limit in an asymptotically flat space are to be expected. Consequently, any approach to treat galactic dynamics as a fully nonlinear problem within general relativity, such as the work discussed by Cooperstock at this conference [56, 57], is to be expected as potentially viable and compatible with the timescape scenario.

As discussed above, in the timescape scenario fits to supernova data, the BAO scale and the angular diameter distance of the sound horizon in CMB anisotropy data allow

one to estimate the dressed matter density parameter, Ω_{M0}, while primordial nucleosynthesis bounds allow us to independently estimate the corresponding baryonic matter density parameter, Ω_{B0}, and consequently of the nonbaryonic matter density parameter $\Omega_{C0} \equiv \Omega_{M0} - \Omega_{B0}$. As a result we find a mass ratio of nonbaryonic dark matter to baryonic matter of $\Omega_{C0}/\Omega_{B0} = 3.1^{+2.5}_{-2.4}$, with uncertainties from supernova data alone, or tighter bounds if constraints based on the angular diameter distance to the sound horizon are imposed. This potentially reduces the relative amount of nonbaryonic matter by a factor of two or more as compared to the standard concordance cosmology.

The Cooperstock–Tieu model [56, 57] demonstrates that the rotation curves of spiral galaxies can be reproduced[9] by a stationary axisymmetric rotating dust solution, obviating the need for spherical halos of dark matter as a required by the naïve use of Newtonian dynamics[10]. However, the Cooperstock–Tieu model does not specify the particle content of the dust. For the Milky Way the Cooperstock–Tieu mass estimate[11] of $2.1 \times 10^{11} M_\odot$ [57] is a factor of 3–5 times *larger* than direct estimates of the combined baryonic mass of the galactic disk, bulge, bar and nucleus which are in the range[12] $(4.2 - 7.2) \times 10^{10} M_\odot$ [61]. The ratio of the Cooperstock–Tieu mass to the observed baryonic mass of the Milky Way is thus in agreement with the global timescape estimate $\Omega_{M0}/\Omega_{B0} = 4.1^{+2.5}_{-2.4}$. Consequently, it is certainly possible for a significant amount of nonbaryonic dark matter to exist within the universe, reduced relative to the Newtonian dynamics estimate, if the Cooperstock–Tieu model, or something close to it[13], operates at the galactic level.

The fact that several modified gravity approaches, including MOND [63], MOG [64], and conformal gravity [36], are able to phenomenologically reproduce various aspects of galactic and galaxy cluster dynamics to varying degrees of success, suggests that some simplifying principle remains to be found despite the amazing variety of structures described, which are far too complex to be modelled by simple exact dust solutions of general relativity such as those of refs. [56, 57]. Given a number of models which fit the same data [65], what is needed is that other falsifiable predictions of all the models

---

[9] See ref. [58] for recent work which increases the number of galaxies whose rotation curves have been successfully fit by this model.

[10] A number of details of the Cooperstock–Tieu model have been disputed; see, e.g., ref. [59] for a summary. While the details are open to debate [57, 59], any deficiencies of the model might easily be an artefact of the simplifications of not including gas pressure, or other realistic features such as differentially rotating spiral arms and central bars. The basic premise of the Cooperstock–Tieu model, namely that nonlinearities in the Einstein equations can be important even in the weak field regime, stands as a consequence of general relativity that must be seriously considered at the galactic level.

[11] A recent Newtonian estimate of the mass of the Milky Way [60], including its dark halo, gives a mass in the range $(5.7 - 10) \times 10^{11} M_\odot$ within a radius 80 kpc, or a total virial mass of $(1.6 \pm 0.3) \times 10^{12} M_\odot$ at the virial radius $R_{\rm vir} = 300$ kpc. A direct numerical comparison with the Cooperstock–Tieu mass estimate is difficult, as the latter is confined to the mass within 30 kpc of the galaxy centre for which rotation curve data was available, and different density profiles can be assumed in the dark outer regions. However, by any measure the Cooperstock–Tieu mass estimate is certainly considerably reduced relative to the Newtonian estimate based on a dark matter halo.

[12] The central supermassive black hole, of mass $\sim 4 \times 10^9 M_\odot$ is not included in this estimate of the Milky Way baryonic mass, as it is omitted in the Cooperstock–Tieu model.

[13] Another variant of the Cooperstock–Tieu model has been explored by Balasin and Grumiller [62].

are developed, which might rule some out. The standard Newtonian CDM hypothesis is at least so well developed that it leads to several testable predictions about structure formation; to the extent that it is arguably *already ruled out* by detailed observational studies of Local Group galaxies [66].

Although naïvely the timescape model appears to predict of order three times as much nonbaryonic matter as baryonic matter by mass, we should be careful to note that the ratio Ω_{M0}/Ω_{B0} measures the density of clumped matter at the present epoch relative to that of baryonic matter inferred from primordial nucleosynthesis, only when using solutions *averaged on cosmological scales*. Given the problems of defining gravitational energy in the absence of a timelike Killing vector or other exact symmetries, we should remain open to the possibility that the nature of clumped gravitational mass in cosmological averages is more than simply the sum of its particle constituents. Thus the difference, $\Omega_{M0} - \Omega_{B0}$, might not simply be nonbaryonic dark matter particles, but could include some component of gravitational energy that enters on some relevant scale of coarse-graining of bound systems, such as the transition from individual galaxies to galaxy clusters. Even for individual galaxies, such the Milky Way, the difference between the Cooperstock–Tieu mass estimate and the observed baryonic mass estimate could either simply be unaccounted dark matter (baryonic or nonbaryonic), or else include at least a partial contribution from gravitational energy that enters in the coarse–graining of the dust. Thus we should keep an open mind about the existence or nonexistence of nonbaryonic dark matter as long as these questions are not understood.

It is worth mentioning, however, that in the standard cosmology nonbaryonic dark matter is required to start structure formation going. At the surface of last scattering, dark matter density contrasts $\delta\rho_C/\rho_C$ are expected to be an order of magnitude stronger than the baryonic density contrasts $\delta\rho_B/\rho_B \sim 10^{-5}$. The nonbaryonic dark matter overdensity contrasts provide the seed gravitational wells into which baryons fall. Although the relative amounts of nonbaryonic dark matter are reduced in the straightforward interpretation of the timescape scenario, there would be likely to be few changes to the basic qualitative scenario of the initiation of standard structure formation. If one wishes to completely eliminate nonbaryonic dark matter on the other hand, then one faces the formidable challenge of explaining how realistic structures can form from density contrasts which are only of order 10^{-5} at last scattering. Thus the timescape scenario with the difference, $\Omega_{M0} - \Omega_{B0}$, interpreted as a nonbaryonic dark matter component remains the simplest scenario from the viewpoint of present understanding.

DISCUSSION

Any serious physical theory should not only be founded on sound principles, but also provide predictions that can potentially rule it out. Much of the present review has therefore concentrated on several tests which might distinguish the timescape model from models of homogeneous dark energy. The (in)homogeneity test of Clarkson, Bassett and Lu is a definitive test independent of the timescape model with the potential to falsify the standard cosmology on large scales, since it tests the validity of the Friedmann equation directly. It would similarly rule out any modified gravity model which relied on a homogeneous geometry with a Friedmann–type equation at the largest scales.

In performing any tests, however, one must be very careful to ensure that data has not been reduced with built–in assumptions that use the Friedmann equation. For example, current estimates of the BAO scale, such as that of Percival *et al*. [46], do not determine D_V directly from redshift and angular diameter measures, but first perform a Fourier space transformation to a power spectrum, assuming a FLRW cosmology. Redoing such an analysis for the timescape model may involve a recalibration of relevant transfer functions.

In the case of supernovae, one must also take care since compilations such as the Union [52], Constitution [53] and Union2 [54] datasets use the SALT or SALT-II methods to calibrate light curves. In this approach empirical light curve parameters and cosmological parameters – *assuming the Friedmann equation* – are simultaneously fit by analytic marginalisation before the raw apparent magnitudes are recalibrated. As Hicken *et al*. discuss [53], a number of systematic discrepancies exist between data reduced by the different methods even within the ΛCDM model. In the case of the timescape model, we find considerable differences between the different approaches [27], which appear to be largely due to systematic issues in distinguishing reddening by host galaxy dust from an intrinsic colour variation in the supernovae. It is also crucial for the timescape scenario that data is cut at the scale of statistical homogeneity ($z \sim 0.033$), below which a simple average Hubble law is not expected. For datasets reduced by the SALT or SALT-II methods there is generally Bayesian evidence that favours the ΛCDM model over the TS model. By contrast for datasets reduced by MLCS2k2 the Bayesian evidence favours the TS model over the ΛCDM model [27]. In principle, with perfect standard candles there are already enough supernovae to decide between the ΛCDM and timescape models on Bayesian evidence, but in practice one is led to different conclusions depending on how the data is reduced. It is therefore important that the systematic issues are unravelled.

The value of the dressed Hubble constant is also an observable quantity of considerable interest. A recent determination of H_0 by Riess *et al*. [55] poses a challenge for the timescape model. However, it is a feature of the timescape model that a 17–22% variance in the apparent Hubble flow will exist on local scales below the scale of statistical homogeneity, and this may potentially complicate calibration of the cosmic distance ladder. Further quantification of the variance in the apparent Hubble flow in relationship to local cosmic structures would provide an interesting possibility for tests of the timescape cosmology for which there are no counterparts in the standard cosmology.

A huge amount of work remains to be done to develop the timescape scenario to the level of detail of the standard cosmology. At the mathematical level, we need to refine the notion of coarse–graining of dust in relation to the various scales of averaging, slicings by hypersurfaces in the evolution equations, and null cone averages. Whatever the outcome of such investigations, I find it exciting that much remains still to be explored in general relativity.

As long as the number of alternative theories is comparable to the number of "alternative" theorists, the detailed development of any alternative paradigm to the standard cosmology may take several years or decades, a timescale which also applies to the big science projects needed to perform precision observations such as redshift-time drift test [49]. Since one can always achieve better fits by adding new terms to relevant equations, every theorist is inevitably guided by intuition and aesthetic judgements about physical

principles as much as by existing observations and experiments.

My own theoretical prejudices are rooted in the knowledge that general relativity is on one hand an extremely successful theory of nature, in complete agreement with observations on the scale of stellar systems, and yet on the other hand, although it is based on deep physical principles, it is still also a theory which has not been completely understood in terms of the coarse-graining of dust, averaging, fitting, the statistical nature of gravitational energy and entropy, and the nature of Mach's principle. Although the nonlinearities of the Einstein equations may play a role in unravelling the mystery of dark matter [56, 57], my own opinion is that what is at stake is more than simply nonlinear mathematics, but also deep and subtle questions of physical principle.

Even if the retro-fit of a density distribution to observed galaxy rotation velocities via Einstein's equations [56, 57] could be independently shown to closely match the observed density distribution, there may be more subtle issues relating to coarse–graining and averaging which underlie the formation of the observed dust distributions, which may also be phenomenologically applicable to galaxies or galaxy clusters with less symmetry. It is worth noting that MOG [64] operates by an effective phenomenological variation of Newton's constant. Since direct observations never directly involve G but rather GM, my suspicion is that the phenomenology is pointing to the thorny issue of the definition of gravitational energy when averaging on different scales. This is the question we need to think more deeply about. The difficult problem of quasi-local gravitational energy in Einstein's theory may turn out not to simply be an arcane curiosity in mathematical relativity, but to be of direct importance for understanding the large scale structure of the universe.

ACKNOWLEDGMENTS

I thank Prof. Remo Ruffini and ICRANet for support and hospitality while the work of refs. [3, 5] was undertaken. This work was also partly supported by the Marsden fund of the Royal Society of New Zealand. I am grateful to many colleagues and students for numerous discussions, including in particular Thomas Buchert, Teppo Mattsson, Roy Kerr and Peter Smale. I also thank John Auping for correspondence which led to the inclusion of the discussion about nonbaryonic dark matter.

REFERENCES

1. A. Einstein, Sitzungsber. Preuss. Akad. Wiss., (1917) 142 [English translation in *"The collected papers of Albert Einstein. Vol.* **6***"* (Princeton Univ. Press, 1997) pp. 421–432].
2. D. L. Wiltshire, *New J. Phys.* **9**, 377 (2007).
3. D. L. Wiltshire, *Phys. Rev.* **D 78**, 084032 (2008).
4. D. L. Wiltshire, *Int. J. Mod. Phys.* **D 18**, 2121 (2009).
5. D. L. Wiltshire, *Phys. Rev.* **D 80**, 123512 (2009).
6. D. L. Wiltshire, *Phys. Rev. Lett.* **99**, 251101 (2007).
7. D. L. Wiltshire, in *Dark Matter in Astroparticle and Particle Physics: Proc. of the 6th International Heidelberg Conference*, eds H. V. Klapdor–Kleingrothaus and G. F. Lewis, (World Scientific, Singapore, 2008) pp. 565-596 [arXiv:0712.3984].
8. F. Hoyle and M. S. Vogeley, *Astrophys. J.* **566**, 641 (2002); *Astrophys. J.* **607**, 751 (2004).

9. K. Bolejko, A. Krasiński, C. Hellaby and M.-N. Célérier, *Structures in the Universe by Exact Methods*, (Cambridge University Press, 2009).
10. G. Lemaître, Ann. Soc. Sci. Bruxelles **A 53**, 51 (1933) [English translation in Gen. Relativ. Grav. **29**, 641 (1997)]; R.C. Tolman, Proc. Nat. Acad. Sci. **20**, 169 (1934); H. Bondi, *Mon. Not. R. Astr. Soc.* **107**, 410 (1947).
11. R. J. van den Hoogen, arXiv:1003.4020.
12. T. Buchert, *Gen. Relativ. Grav.* **32**, 105 (2000); *Gen. Relativ. Grav.* **33**, 1381 (2001).
13. T. Buchert, *Gen. Relativ. Grav.* **40**, 467 (2008).
14. P. J. E. Peebles, *AIP Conf. Proc.* **1241**, 175 (2010).
15. L.B. Szabados, Living Rev. Rel. **7**, 4 (2004).
16. H. Bondi, *Cosmology*, (Cambridge Univ. Press, 1961).
17. J. Bičák, J. Katz and D. Lynden-Bell, *Phys. Rev.* **D 76**, 063501 (2007).
18. S. Räsänen, *JCAP* 03, 018 (2010).
19. K. Bolejko and L. Andersson, *JCAP* 10, 003 (2008).
20. R. A. Sussman, arXiv:0807.1145; arXiv:0809.3314; *Class. Quantum Grav.* **27**, 175001 (2010).
21. M. Mattsson and T. Mattsson, *JCAP* 10, 021 (2010); arXiv:1012.4008.
22. M. Korzynski, *Class. Quantum Grav.* **27**, 105015 (2010).
23. A. Wiegand and T. Buchert, *Phys. Rev.* **D 82**, 023523 (2010).
24. T. Buchert, in this volume [arXiv:1012.3084].
25. B. M. Leith, S. C. C. Ng and D. L. Wiltshire, *Astrophys. J.* **672**, L91 (2008).
26. A. G. Riess *et al*., *Astrophys. J.* **659**, 98 (2007).
27. P. R. Smale and D. L. Wiltshire, arXiv:1009.5855, to appear in *Mon. Not. R. Astr. Soc.* (2011).
28. S. Jha, A. G. Riess and R. P. Kirshner, *Astrophys. J.* **659**, 122 (2007).
29. N. Li and D. J. Schwarz, *Phys. Rev.* **D 78**, 083531 (2008).
30. R. Watkins, H.A. Feldman and M.J. Hudson, 2009, *Mon. Not. R. Astr. Soc.* **392**, 743.
31. A. Kashlinsky, F. Atrio-Barandela, D. Kocevski and H. Ebeling, *Astrophys. J.* **686**, L49 (2009); A. Kashlinsky, F. Atrio-Barandela, H. Ebeling, A. Edge and D. Kocevski, *Astrophys. J.* **712**, L81 (2010).
32. R.B. Tully, E.J. Shaya, I.D. Karachentsev, H. Courtois, D.D. Kocevski, L. Rizzi and A. Peel, *Astrophys. J.* **676**, 184 (2008).
33. E. Komatsu *et al*., *Astrophys. J. Suppl.* **180**, 330 (2009).
34. B. E. Schaefer, *Astrophys. J.* **660**, 16 (2007); N. Liang, W. K. Xiao, Y. Liu and S. N. Zhang, *Astrophys. J.* **685**, 354 (2008); L. Amati, C. Guidorzi, F. Frontera, M. Della Valle, F. Finelli, R. Landi and E. Montanari, *Mon. Not. R. Astr. Soc.* **391**, 577 (2008); R. Tsutsui, T. Nakamura, D. Yonetoku, T. Murakami, Y. Kodama and K. Takahashi, *JCAP* 08, 015 (2009).
35. B. E. Schaefer, talk at *Sixth Huntsville Gamma Ray Burst Symposium 2008*, http://grbhuntsville2008.cspar.uah.edu/content/Talks/Schaefer.pdf
36. P. D. Mannheim, *Prog. Part. Nucl. Phys.* **56**, 340 (2006).
37. G. B. Zhao and X. Zhang, *Phys. Rev.* **D 81**, 043518 (2010).
38. P. Serra, A. Cooray, D. E. Holz, A. Melchiorri, S. Pandolfi and D. Sarkar, *Phys. Rev.* **D 80**, 121302 (2009).
39. E. Gaztañaga, A. Cabré and L. Hui, *Mon. Not. R. Astr. Soc.* **399**, 1663 (2009).
40. J. A. Gu, C. W. Chen and P. Chen, *New J. Phys.* **11**, 073029 (2009).
41. V. Sahni, A. Shafieloo and A. A. Starobinsky, *Phys. Rev.* **D 78**, 103502 (2008).
42. C. Zunckel and C. Clarkson, *Phys. Rev. Lett.* **101**, 181301 (2008).
43. A. Shafieloo, V. Sahni and A. A. Starobinsky, *Phys. Rev.* **D 80**, 101301 (2009).
44. C. Alcock and B. Paczyński, *Nature* **281**, 358 (1979).
45. D. J. Eisenstein *et al*., *Astrophys. J.* **633**, 560 (2005); S. Cole *et al*., *Mon. Not. R. Astr. Soc.* **362**, 505 (2005).
46. W. J. Percival *et al*., *Mon. Not. R. Astr. Soc.* **381**, 1053 (2007); W. J. Percival *et al*., *Mon. Not. R. Astr. Soc.* **401**, 2148 (2010).
47. E. Gaztañaga, A. Cabré, F. Castander, M. Crocce and P. Fosalba, *Mon. Not. R. Astr. Soc.* **399**, 801 (2009).
48. C. Clarkson, B. Bassett and T. C. Lu, *Phys. Rev. Lett.* **101**, 011301 (2008).
49. A. Sandage, *Astrophys. J.* **136**, 319 (1962); G. C. McVittie, *Astrophys. J.* **136**, 334 (1962); A. Loeb, *Astrophys. J.* **499**, L111 (1998).

50. J. P. Uzan, C. Clarkson and G. F. R. Ellis, *Phys. Rev. Lett.* **100**, 191303 (2008).
51. P. S. Corasaniti, D. Huterer and A. Melchiorri, *Phys. Rev.* **D 75**, 062001 (2007); J. Liske *et al.*, *Mon. Not. R. Astr. Soc.* **386**, 1192 (2008).
52. M. Kowalski *et al*., *Astrophys. J.* **686**, 749 (2008).
53. M. Hicken *et al*., *Astrophys. J.* **700**, 1097 (2009).
54. R. Amanullah *et al*., *Astrophys. J.* **716**, 712 (2010).
55. A. G. Riess *et al*., *Astrophys. J.* **699**, 539 (2009).
56. F. I. Cooperstock and S. Tieu, *Mod. Phys. Lett.* **A 21**, 2133 (2006).
57. F. I. Cooperstock and S. Tieu, *Int. J. Mod. Phys.* **A 22**, 2293 (2007).
58. J. D. Carrick and F. I. Cooperstock, arXiv:1101.3224.
59. A. Rakić and D. J. Schwarz, *PoS* **IDM2008**, 096 (2008) [arXiv:0811.1478].
60. O. Y. Gnedin, W. R. Brown, M. J. Geller and S. J. Kenyon, *Astrophys. J.* **720**, L108 (2010).
61. A. Klypin, H. Zhao and R. S. Somerville, *Astrophys. J.* **573**, 597 (2002).
62. H. Balasin and D. Grumiller, *Int. J. Mod. Phys.* **D 17**, 475 (2008).
63. M. Milgrom, *Astrophys. J.* **270**, 365 (1983).
64. J. W. Moffat, *JCAP* 03, 004 (2006).
65. J. R. Brownstein, Ph.D. thesis (University of Waterloo, 2009) [arXiv:0908.0040].
66. P. Kroupa *et al*., *Astron. Astrophys.* **523**, A32 (2010).

Cosmological consequences of Modified Gravity (MOG)

Viktor T. Toth

Ottawa, ON K1N 9H5, Canada

Abstract. As an alternative to the ΛCDM concordance model, Scalar-Tensor-Vector Modified Gravity (MOG) theory reproduces key cosmological observations without postulating the presence of an exotic dark matter component. MOG is a field theory based on an action principle, with a variable gravitational constant and a repulsive vector field with variable range. MOG yields a phenomenological acceleration law that includes strong tensorial gravity partially canceled by a repulsive massive vector force. This acceleration law can be used to model the CMB acoustic spectrum and the matter power spectrum yielding good agreement with observation. A key prediction of MOG is the presence of strong baryonic oscillations, which will be detectable by future surveys. MOG is also consistent with Type Ia supernova data. We also describe on-going research of the coupling between MOG and continuous matter, consistent with the weak equivalence principle and solar system observations.

Keywords: Cosmology, modified gravity, CMB acoustic spectrum, matter power spectrum, cosmic acceleration
PACS: 04.20.Cv,04.50.Kd,04.80.Cc,45.20.D-,45.50.-j,98.80.-k

1. INTRODUCTION

Why is there a need for a modified gravity theory? There is a perfectly serviceable model of cosmology, the so-called ΛCDM "concordance" model, that is not only in good agreement with a large body of observational evidence, it also yielded some impressive predictions. Nonetheless, we feel motivated to seek alternatives, in part for the following reasons:

- The ΛCDM model requires 96% of the universe to consist of black "stuff": cold dark matter and dark energy, both of which may never be detectable except through their gravitational influence;
- The cold dark matter paradigm runs into difficulties even closer to home, notably its inability to explain convincingly why the rotation curves of spiral galaxies so closely follow their luminosity profiles.

The modified gravity theory we discuss here, Scalar-Tensor-Vector Gravity [1, 2] (STVG), has also been referred to by the acronym MOG more recently. MOG is a particularly interesting candidate for gravity modification in part because:

- In the solar system or the laboratory, MOG predicts Newtonian (or Einsteinian) physics;
- The MOG acceleration law is consistent with star clusters [3], galaxies [4], and galaxy clusters [5, 6].

In the rest of this presentation, we show that MOG also appears to be consistent with cosmological data [7, 8]. If these results hold, MOG may prove to be a more economical theory (in the sense of Occam's razor) than ΛCDM.

We begin with introducing MOG as a Lagrangian field theory in Section 2, also discussing the subject of coupling between the MOG fields and matter. Next, we briefly introduce the phenomenology of MOG, concentrating mostly on the spherically symmetric, static vacuum solution in Section 3. We then move on to cosmology: we discuss the MOG prediction of the acoustic spectrum of the Cosmic Microwave Background (CMB) in Section 4 and the galaxy-galaxy matter power spectrum in Section 5. Finally, we move on to the topic that is the most active area of our current research, the study of MOG in the presence of continuous matter such as a perfect fluid, in Section 6. We conclude with a brief discussion of the most significant challenges and outlook in Section 7.

2. MOG AS A FIELD THEORY

MOG is a theory of gravity that augments Einstein's gravitational theory with a variable gravitational constant and a massive vector field with variable mass and coupling strength, producing a repulsive force. The theory's building blocks are:

- The tensor field $g_{\mu\nu}$ of metric gravity;
- A scalar field G representing a variable gravitational constant;
- A massive vector field ϕ_μ responsible for a repulsive force;
- Another scalar field μ representing the variable mass of the vector field;
- A further scalar field ω representing the variable coupling strength of the vector field.[1]

MOG is a theory based on a Lagrangian action principle. The MOG Lagrangian has three parts: the Einstein-Hilbert Lagrangian of tensor gravity, the Lagrangian of the massive vector field, and the Lagrangian of the three scalar fields, complete with self-interaction potentials:

$$\begin{aligned}\mathscr{L} =& -\frac{1}{16\pi G}\left(R+2\Lambda\right)\sqrt{-g} \\ & -\frac{1}{4\pi}\omega\left[\frac{1}{4}B^{\mu\nu}B_{\mu\nu}-\frac{1}{2}\mu^2\phi_\mu\phi^\mu+V_\phi(\phi)\right]\sqrt{-g} \\ & -\frac{1}{G}\left[\frac{1}{2}g^{\mu\nu}\left(\frac{\nabla_\mu G\nabla_\nu G}{G^2}+\frac{\nabla_\mu\mu\nabla_\nu\mu}{\mu^2}-\nabla_\mu\omega\nabla_\nu\omega\right)+\frac{V_G(G)}{G^2}+\frac{V_\mu(\mu)}{\mu^2}-V_\omega(\omega)\right]\sqrt{-g}.\end{aligned} \tag{1}$$

Here, $B_{\mu\nu}=\partial_\mu\phi_\nu-\partial_\nu\phi_\mu$, and $V_\phi(\phi)$, $V_G(G)$, $V_\omega(\omega)$, and $V_\mu(\mu)$ denote the self-interaction potentials associated with the vector field and the three scalar fields. The symbol ∇_μ is used to denote covariant differentiation with respect to the metric $g^{\mu\nu}$, while the symbols R, Λ, and g represent the Ricci-scalar, the cosmological constant,

[1] Although ω is included for generality, in the solutions that we studied it turns out to be a constant.

and the determinant of the metric tensor, respectively. We define the Ricci tensor as $R_{\mu\nu} = \partial_\alpha \Gamma^\alpha_{\mu\nu} - \partial_\nu \Gamma^\alpha_{\mu\alpha} + \Gamma^\alpha_{\mu\nu}\Gamma^\beta_{\alpha\beta} - \Gamma^\alpha_{\mu\beta}\Gamma^\beta_{\alpha\nu}$. Our units are such that the speed of light, $c = 1$; we use the metric signature $(+,-,-,-)$.

The vector field is expected to produce a repulsive force. This is not possible unless matter carries a vector charge. Furthermore, the vector charge must have the right strength to cancel out excess gravity exactly such that the effective gravitational constant that remains is that of Newton. This means that the coupling term must also include a dependence on the scalar field G. This is important for another reason as well: a scalar charge is required in order to ensure that the theory survives precision solar system tests [9].

We specify this coupling in the case of a massive test particle by explicitly incorporating it into the test particle Lagrangian:

$$\mathscr{L}_{\rm TP} = -m + \alpha\omega q_5 \phi_\mu u^\mu, \tag{2}$$

where α is a function of G and q_5 is the vector charge of a test particle with mass m and four-velocity u^μ.

This Lagrangian has been used in conjunction with the spherically symmetric, static vacuum solution of the MOG field equations to derive the phenomenology that we discuss in the next section.

3. MOG PHENOMENOLOGY

In MOG, the metric tensor is responsible for Einstein-like gravity, but G is generally greater than Newton's constant, G_N.

The vector field is responsible for a repulsive force, canceling out part of the gravitational force; the effective gravitational constant at short range is G_N. The vector field is massive and has limited range; beyond its range, gravity is stronger than Newton predicts.

The strength of G and the range μ^{-1} of the vector field are determined by the source mass.

In the weak field, low velocity limit, the acceleration due to a spherically symmetric source of mass M is

$$\ddot{r} = -\frac{G_N M}{r^2}\left[1 + \alpha - \alpha(1+\mu r)e^{-\mu r}\right], \tag{3}$$

where the overdot denotes differentiation with respect to time. The values of α and μ are determined by the source mass M with formulas fitted using galaxy rotation and cosmology data:

$$\alpha = \frac{M}{(\sqrt{M}+E)^2}\left(\frac{G_\infty}{G_N} - 1\right), \qquad \mu = \frac{D}{\sqrt{M}}, \tag{4}$$

$$D \simeq 6250\, M_\odot^{1/2}\ \text{kpc}^{-1}, \qquad E \simeq 25000\, M_\odot^{1/2}, \qquad G_\infty = (1+\alpha)G_N \simeq 20 G_N.$$

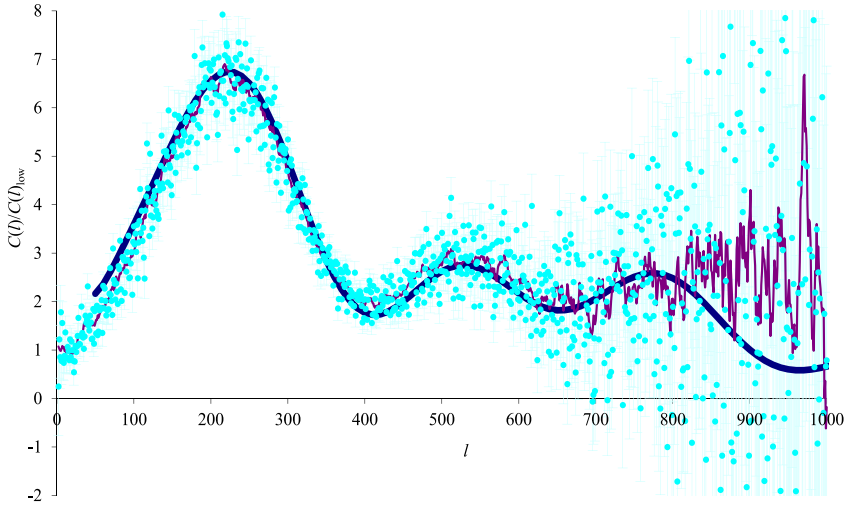


FIGURE 1. The acoustic spectrum of the cosmic microwave background (5-year WMAP data points [10] with error bars in light blue, with a moving window average also shown in purple) and the ΛCDM prediction (thick blue line).

This acceleration law is consistent with laboratory and solar system experiments, star clusters, galaxies, and galaxy clusters across (at least) 15 orders of magnitude. At short range, $\mu r \ll 1$, (3) becomes Newton's gravitational acceleration law,

$$\ddot{r} \simeq -\frac{G_N M}{r^2}, \tag{5}$$

whereas at great distances, $\mu r \gg 1$, we get Newtonian gravity with an "enhanced" value of the gravitational constant,

$$\ddot{r} \simeq -(1+\alpha)\frac{G_N M}{r^2}. \tag{6}$$

We also used this acceleration law to investigate the MOG predictions for the cosmic microwave background and the galaxy-galaxy matter power spectrum, which we discuss below.

4. MOG AND THE CMB

One of the key successes of the standard model of cosmology, ΛCDM, is its ability to predict the position and size of peaks in the acoustic power spectrum of the cosmic microwave background (Figure 1).

The question naturally arises: can MOG reproduce this result, especially in view of the fact that there exists no exotic dark matter in the MOG cosmology?

As we were attempting to address this question, colleagues often advised us to use the "industry standard" cosmological code `CMBFAST` [11] or one of its derivatives such as `CMBEASY` [12]. When we initiated a detailed study of these programs, however, we found that adapting them to a variable-G cosmology is a highly nontrivial undertaking. A key reason is that CMBFAST uses variants of the cosmological quantity Ω (e.g., the

```
      zeqp1=2.5d4*omegam*h*h*(2.7d0/tcmb)**4
c     EH (97) fitting formula for masive neutrino growth factor.
      fnu=omegan/omegam
      fcb=(omegac+omegab)/omegam
      if (fnu.gt.0.0d0) then
         apcb=0.25d0*(5.0d0-sqrt(1.0d0+24.0d0*fcb))
         aktoq=(2.7d0/tcmb)*(2.7d0/tcmb)/(omegam*h*h)*h
         yfsok2=17.2d0*fnu*(1.0d0+0.488d0
     &        *exp(-7.0d0*log(fnu)/6.0d0))
     &        *(dble(annunr)*aktoq/fnu)**2
      else
         apcb=1.0d0
         yfsok2=0.0d0
      end if
```

FIGURE 2. Code fragment from the CMBFAST [11] source distribution, demonstrating the difficulty of applying the program to a variable-G cosmology, due to the fact that $\Omega \propto G\rho$ is used to represent constituent densities in both gravitational and non-gravitational contexts.

baryon density Ω_b) to represent matter in both gravitational (e.g., structure growth) and nongravitational (e.g., speed-of-sound calculations) contexts. In a variable-G cosmology, $\Omega = 8\pi G\rho/3H^2$ may change even as the corresponding density ρ remains constant, due to changes in the value of G. Whereas gravitational relations involve the quantity $G\rho$, which Ω properly represents, nongravitational relations involve ρ.

This difficulty is by no means insurmountable, but it turns the adaptation of CMBFAST into an arduous and error-prone task.

Other codes, such as CMBEASY, often use a version of CMBFAST as the underlying computational engine. Worse yet, the engine is often machine-translated from the original FORTRAN into another programming language, such as C or C++.

It was in part for this reason that we elected to take a closer look at a promising alternative: a semi-analytical approximation[2] developed by Mukhanov [14] that is nonetheless more than just a collection of fitting formulae. Mukhanov's formulation does not hide the underlying physics, and it becomes a relatively straightforward substitution to replace, e.g., all occurrences of Ω with $(G_{\rm eff}/G_N)\Omega$ in gravitational contexts, where $G_{\rm eff}$ is the effective gravitational constant at the horizon. The result (Figure 3) is encouraging but not altogether surprising: The enhanced gravitational constant plays the same role as collisionless dark matter in structure growth, but dissipation is due to the baryonic matter density, which is the same as in ΛCDM.

5. MOG AND THE MATTER POWER SPECTRUM

Another key prediction of ΛCDM cosmology, confirmed by observation, is the matter power spectrum: the statistic of density fluctuations in the large-scale distribution of galaxies.

[2] It should be noted that the CMBFAST code base also relies on semi-analytic formulations, e.g., those developed by Eisenstein and Hu [13].

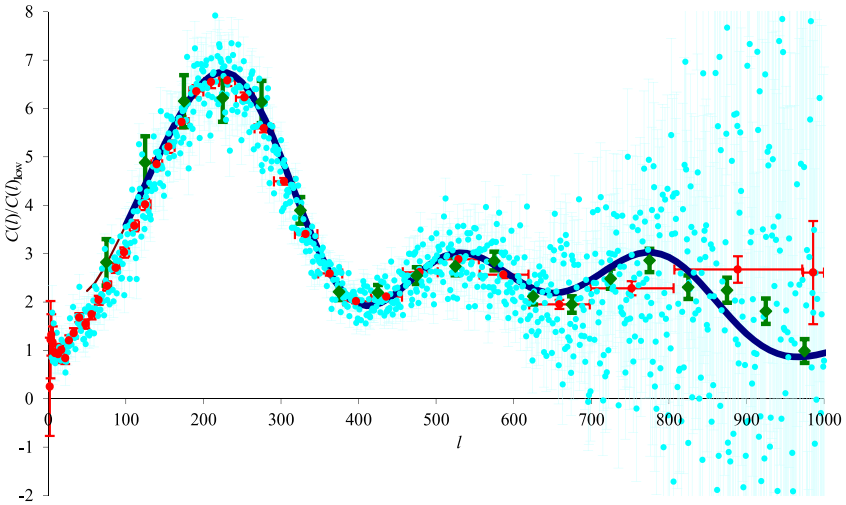


FIGURE 3. The acoustic spectrum of the cosmic microwave background (WMAP data points with error bars in light blue) and the MOG prediction (thick blue line). Binned WMAP data (red) and Boomerang data (green) are also shown.

This statistic can be understood using the Newtonian theory of small fluctuations in a self-gravitating medium. These fluctuations are governed by the equation

$$\ddot{\delta}_{\mathbf{k}} + \frac{\dot{a}}{a}\dot{\delta}_{\mathbf{k}} + \left(\frac{C_s^2 k^2}{a^2} - 4\pi G\rho\right)\delta_{\mathbf{k}} = 0 \tag{7}$$

for each Fourier mode $\delta = \delta_{\mathbf{k}}(t)e^{i\mathbf{k}\cdot\mathbf{q}}$ (such that $\nabla^2\delta = -k^2\delta$; C_s is the speed of sound in the medium). In the case of the standard model cosmology, ∇^2 can be obtained from the Poisson equation of Newtonian gravity. In MOG, the acceleration law can be used to derive the corresponding inhomogeneous Helmholtz equation:

$$\nabla^2\Phi = 4\pi G_N\rho(\mathbf{r}) + \alpha\mu^2 G_N\int\frac{e^{-\mu|\mathbf{r}-\tilde{\mathbf{r}}|}\rho(\tilde{\mathbf{r}})}{|\mathbf{r}-\tilde{\mathbf{r}}|}\,d^3\tilde{\mathbf{r}}. \tag{8}$$

The Helmholtz equation leads a shifting of the wave number in the solution to (7):

$$k'^2 = k^2 + 4\pi a^2\left(\frac{G_{\rm eff}-G_N}{G_N}\right)\lambda_J^{-2}. \tag{9}$$

Changes to the sound horizon scale are unaffected by the varying strength of gravity. However, Silk damping introduces a $G^{3/4}$ dependence [15]:

$$k'_{\rm Silk} = k_{\rm Silk}\left(\frac{G_{\rm eff}}{G_N}\right)^{3/4}. \tag{10}$$

These results can be substituted in the analytical approximations of Eisenstein and Hu [13], leading to the the plot shown in Figure 4.

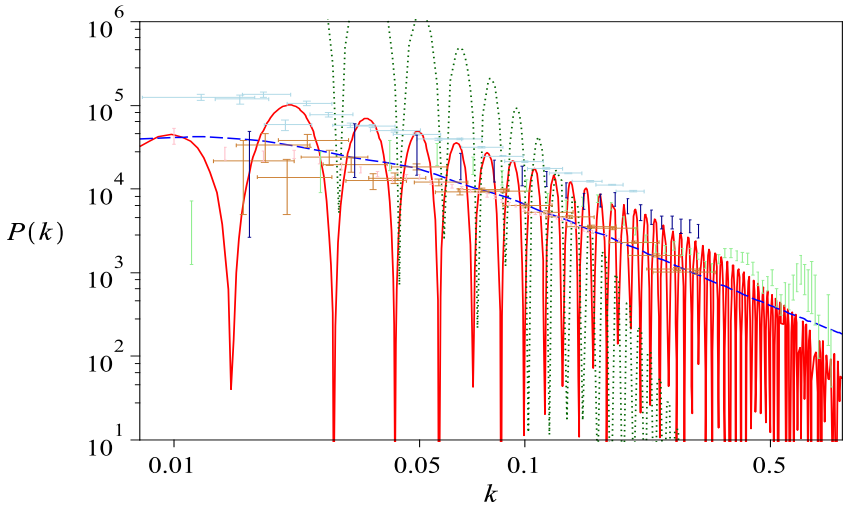


FIGURE 4. The matter power spectrum. Three models are compared against five data sets: ΛCDM (dashed blue line, $\Omega_b = 0.035$, $\Omega_c = 0.245$, $\Omega_\Lambda = 0.72$, $H = 71$ km/s/Mpc), a baryon-only model (dotted green line, $\Omega_b = 0.035$, $H = 71$ km/s/Mpc), and MOG (solid red line, $\alpha = 19$, $\mu = 5h$ Mpc^{-1}, $\Omega_b = 0.035$, $H = 71$ km/s/Mpc.) Data points are colored light blue (SDSS 2006 [16]), gold (SDSS 2004 [17]), pink (2dF [18]), light green (UKST [19]), and dark blue (CfA [20]).

Comparing the result visually against the data, we can tell that the solution appears to have the right slope; however, unlike the ΛCDM prediction, MOG predicts unit oscillations in the power spectrum. These unit oscillations are dampened in the case of ΛCDM by the presence of matter; in the case of MOG, no such matter is present and oscillations are not dampened.

Are these unit oscillations present in the data? As it turns out, it is not yet possible to answer this question as the resolution of the data set is not sufficient. The data are effectively binned using a window function. When we apply that window function to the MOG prediction, the unit oscillations are smoothed out, and the result shows very good agreement indeed with the data points (Figure 5).

In summary, two key features of the matter power spectrum are its slope and the presence or absence of baryonic oscillations. MOG reproduces the right slope; however, unlike ΛCDM, MOG has unit oscillations that are not dampened by the presence of collisionless dark matter. Future galaxy surveys will unambiguously show if unit oscillations are present in the data. Therefore, the matter power spectrum can be key to distinguish modified gravity without exotic dark matter from cold dark matter theories.

6. MOG AND CONTINUOUS MATTER

In the preceding sections, we discussed how MOG can be used to reproduce the observed characteristics of the CMB acoustic spectrum and the matter power spectrum. These results, however, were based on the MOG point particle solution (the point particle equation of motion in the presence of a spherically symmetric, static, vacuum gravitational field.)

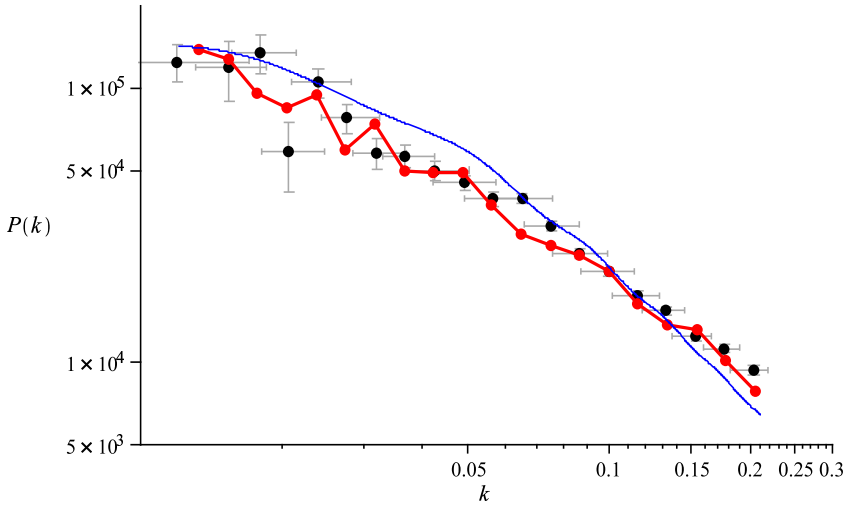


FIGURE 5. The effect of window functions on the power spectrum is demonstrated by applying the SDSS luminous red galaxy survey window functions to the MOG prediction. Baryonic oscillations are greatly dampened in the resulting curve (solid red line). A normalized linear ΛCDM estimate is also shown (thin blue line) for comparison.

The question naturally arises: Is it really appropriate to use the point particle solution for continuous distributions of matter? To answer this question, we have to consider an essential feature of MOG: namely that the strength of the (Newtonian) gravitational field and the range of the MOG repulsive vector force both vary as functions of the source mass. In other words, if we combine two masses together, the resulting field will not be a simple sum of the individual fields, not even in a crude approximation: the theory has an essential nonlinearity.

This means that we cannot blindly rely on the point particle solution to do MOG cosmology: it is necessary to solve the MOG field equations in the presence of continuous matter, such as a (perfect) fluid model. We have taken some tentative steps in this direction (see also [8]).

How does MOG couple to continuous matter? The coupling must be consistent with two constraints:

- MOG must obey the weak equivalence principle (WEP) in order to not run afoul of many observations, e.g., Eötvös-style experiments;
- MOG must be compatible with precision solar system observations, expressed in the form of the parameters of the Parameterized Post-Newtonian (PPN) formalism, notably the Eddington-parameters β and γ.

The two Eddington parameters β and γ determine deviations from the Newtonian potential in post-Newtonian models:

$$g_{00} = 1 - \frac{2M}{r} + 2\beta\left(\frac{M}{r}\right)^2, \tag{11}$$

$$g_{0j} = 0, \tag{12}$$

$$g_{jk} = -\left(1 + 2\frac{2\gamma M}{r}\right)\delta_{jk} \tag{13}$$

The Eddington-parameter β is identically 1 for MOG. The Eddington-parameter γ has the same value as in Jordan-Brans-Dicke theory [21], which can be "cured" by introducing a scalar charge that makes it conformally equivalent to the minimally coupled scalar theory [9].

The WEP is often interpreted as a requirement for a metric theory of gravity. This criteria is obviously not satisfied by MOG, as material particles are assumed to carry a vector charge and not move along geodesics determined by the metric. However, it is possible to consider a more relaxed interpretation of the WEP: a theory that satisfies the WEP observationally must be conformally equivalent to a metric theory of gravity. That is to say that there must exist a conformal transformation under which any non-minimal couplings between matter and non-metric gravity fields would vanish. The justification for this relaxed interpretation is that an observer, equipped with a classical instrument, would not be able to choose between conformally equivalent frames of reference: non-gravitational laws of physics, notably electromagnetism, are unaffected by a conformal transformation.

Conformal transformations add a vector degree of freedom (the special conformal transformation, a translation preceded and followed by an inversion, $x'^{\mu} = (x^{\mu} - b^{\mu}x^2)/(2 - 2b\cdot x + b^2x^2)$) and a scalar degree of freedom (dilation, $x'^{\mu} = \alpha x^{\mu}$); this agrees with the degrees of freedom to which the matter Lagrangian is expected to couple in the MOG theory. The metric tensor is conformally invariant up to a rescaling: $g'^{\mu\nu} = \alpha^{-2}(1 - 2b\cdot x + b^2x^2)^2 g^{\mu\nu}$.

These considerations about the WEP and γ can lead to a tentative general prescription for the coupling between the MOG fields and matter. We anticipate that the field equations for a perfect fluid will contain a vector charge in the form

$$\phi^{\nu} u_{\nu} J_{\mu} = \omega \frac{G - G_N}{G} T_{\mu\nu} u^{\nu}, \tag{14}$$

and a scalar charge in the form

$$GJ = -\frac{1}{2}T. \tag{15}$$

Given an equation of state, we can now write down the MOG field equations in the case of the FLRW metric,

$$ds^2 = dt^2 - a^2(t)\left[(1 - kr^2)^{-1}dr^2 + r^2 d\Omega^2\right]. \tag{16}$$

The equations are, after setting $\omega = \text{const.}$,

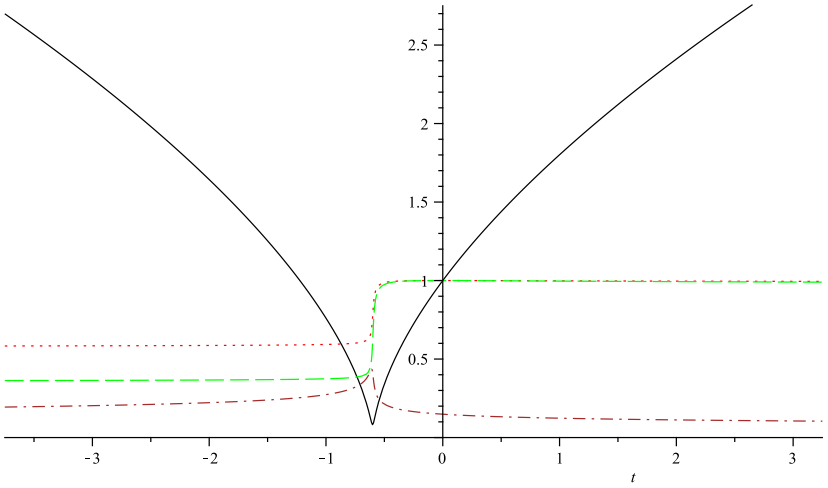


FIGURE 6. The MOG "classical bounce". Black (solid) line is a/a_0 ; red (dotted) line is G/G_0 ; green (dashed) line is μ/μ_0 ; brown (dash-dot) line is $(a^3\rho)/(a_0^3\rho_0)$. Horizontal axis is in units of 13.7 billion years.

$$\left(\frac{\dot{a}}{a}\right)^2+\frac{k}{a^2}=\frac{8\pi G\rho}{3}-\frac{4\pi}{3}\left(\frac{\dot{G}^2}{G^2}+\frac{\dot{\mu}^2}{\mu^2}-\frac{1}{4\pi}G\omega\mu^2\phi_0^2\right)+\frac{2}{3}\omega GV_\phi+\frac{8\pi}{3}\left(\frac{V_G}{G^2}+\frac{V_\mu}{\mu^2}\right)+\frac{\Lambda}{3}+\frac{\dot{a}}{a}\frac{\dot{G}}{G}, \tag{17}$$

$$\frac{\ddot{a}}{a}=-\frac{4\pi G}{3}(\rho+3p)+\frac{8\pi}{3}\left(\frac{\dot{G}^2}{G^2}+\frac{\dot{\mu}^2}{\mu^2}-\frac{1}{4\pi}G\omega\mu^2\phi_0^2\right)+\frac{2}{3}\omega GV_\phi+\frac{8\pi}{3}\left(\frac{V_G}{G^2}+\frac{V_\mu}{\mu^2}\right)+\frac{\Lambda}{3}+H\frac{\dot{G}}{2G}+\frac{\ddot{G}}{2G}-\frac{\dot{G}^2}{G^2}, \tag{18}$$

$$\ddot{G}+3\frac{\dot{a}}{a}\dot{G}-\frac{3}{2}\frac{\dot{G}^2}{G}+\frac{G}{2}\left(\frac{\dot{\mu}^2}{\mu^2}\right)+\frac{3}{G}V_G-V_G'+G\left[\frac{V_\mu}{\mu^2}\right]+\frac{G}{8\pi}\Lambda-\frac{3G}{8\pi}\left(\frac{\ddot{a}}{a}+H^2\right)=-\frac{1}{2}G^2(\rho+3p), \tag{19}$$

$$\ddot{\mu}+3\frac{\dot{a}}{a}\dot{\mu}-\frac{\dot{\mu}^2}{\mu}-\frac{\dot{G}}{G}\dot{\mu}+\frac{1}{4\pi}G\omega\mu^3\phi_0^2+\frac{2}{\mu}V_\mu-V_\mu'=0, \tag{20}$$

$$\omega\mu^2\phi_0-\omega\frac{\partial V_\phi}{\partial\phi_0}=4\pi J_0. \tag{21}$$

These FLRW field equations can be solved numerically, given suitable initial conditions and some assumptions. We generally ignore the self-interaction potentials:

$$V_\phi=V_G=V_\mu=0. \tag{22}$$

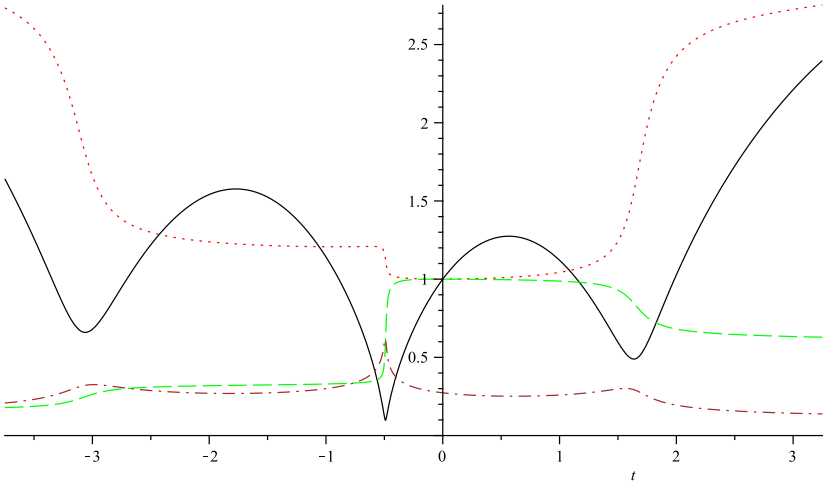


FIGURE 7. The MOG "cyclical bounce" cosmology with negative $V_G = \text{const}$. (For legend, see Figure 6.)

We set the cosmological constant and the curvature constant to zero:

$$\Lambda = 0, \tag{23}$$

$$k = 0. \tag{24}$$

We assume a simple equation of state, $p = w\rho$, and we are mainly interested in the late "dust" universe, $w = 0$.

We use the present epoch to establish initial conditions: e.g., $\dot{a}/a|_{t=t_0} \simeq 2.3 \times 10^{-18}$ s^{-1}, and $\rho|_{t=t_0} \simeq 10^{-26}$ kg/m^3. The solution yields a "classical bounce", albeit with an age problem (Figure 6).

To address the age problem, we can consider using a non-zero value of V_G. We find that a negative value produces a cyclical universe, with repeated classical bounces (Figure 7). The conditions at the present epoch (notably, the expansion rate and approximate density of the universe) are repeated at later times during subsequent cycles. This allows for the possibility of a much older universe, one that has been through several cycles of shrinking and expansion since its densest state.

What about the deceleration parameter? This parameter, defined as $q = -\ddot{a}a/\dot{a}^2$, characterizes the rate at which the expansion rate slows. As it is well known, observations of the luminosity-distance relationship of distant Type Ia supernovae are inconsistent with a flat $q = 0.5$ Einstein-de Sitter universe. The data may be consistent with an empty universe ($q = 0$). It is also consistent with the ΛCDM universe that is dominated by dark energy ($\Omega_\Lambda \simeq 0.7$) at the present epoch.

MOG can also produce good agreement with the Type Ia supernova data if we assume the existence of a small positive value of V_G. This universe has a shallow bounce, and at the present epoch, its evolution is such that $\dot{a}/a$ is nearly constant. The actual age of the universe is, therefore, not fixed by the observed expansion rate alone; fitting to the supernova data yields a universe that is significantly older than the "canonical" value

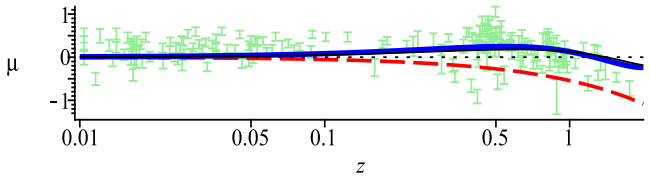


FIGURE 8. Type Ia supernova luminosity-redshift data [22] and the MOG/ΛCDM predictions. The horizontal axis corresponds to the $q = 0$ empty universe. The MOG result is represented by a thick (blue) line. Dashed (red) line is a matter-dominated Einstein de-Sitter universe with $\Omega_M = 1$, $q = 0.5$. Thin (black) line is the ΛCDM prediction.

of 13.7 billion years. This model offers excellent agreement with the Type Ia supernova observations (Figure 8).

All of these models have shortcomings, but their existence shows that MOG is capable of producing physically plausible models of expansion, and that a classical bounce occurs naturally within the theory.

7. CONCLUSIONS AND OUTLOOK

The research of the cosmological consequences of MOG that is presented here is ongoing. Important cosmological results, such as the matter power spectrum and the CMB acoustic spectrum, have been reproduced using the MOG equations of motion in the gravitational field of a point source. A general prescription that describes how MOG couples to matter (while remaining consistent with the WEP and precision observations) will allow us to solve the MOG field equations in the presence of matter [23], allowing us to re-derive the matter power spectrum and CMB acoustic spectrum results using a more solid foundation.

What can MOG tell us about the early universe? A bouncing cosmology may naturally avoid the horizon problem, which is a major motivation for inflationary cosmologies. It is unclear how MOG would address the flatness problem, other than postulating $k = 0$ *a priori*.

What about Big Bang nucleosynthesis? At short range, the MOG acceleration law is consistent with Newton's. However, primordial isotope ratios are sensitive to the rate of cosmic expansion, which is governed by long-range gravity. We do not presently know if MOG can be consistent with observed isotope abundances.

Some of our results were obtained by assuming a constant value for V_G, the self-interaction potential for the G scalar field. This assumption is *ad hoc*, not justified by theory. Perhaps improved solutions can lead us to eliminate the need to postulate a non-zero V_G, or alternatively, a theoretical justification for a (near) constant self-interaction potential can be found.

Notwithstanding these problems and open issues, MOG appears to be consistent with a range of astrophysical and cosmological phenomena. The theory yields a phenomenological acceleration law that works well across some 15 orders of magnitude in scale. Furthermore, there exists an unambiguous cosmological test through which MOG can

be falsified: the absence of baryonic oscillations in the matter power spectrum cannot be explained by a MOG cosmology that lacks a cold dark matter component.

REFERENCES

1. J. W. Moffat, *Journal of Cosmology and Astroparticle Physics* **2006**, 004 (2006), {arXiv:gr-qc/0506021}.
2. J. W. Moffat, and V. T. Toth, *Class. Quant. Grav.* **26**, 085002 (2009), {arXiv:0712.1796[gr-qc]}.
3. J. W. Moffat, and V. T. Toth, *Astrophys. J.* **680**, 1158 (2008), {arXiv:0708.1935[astro-ph]}.
4. J. R. Brownstein, and J. W. Moffat, *Astrophys. J.* **636**, 721–741 (2006), {arXiv:astro-ph/0506370}.
5. J. R. Brownstein, and J. W. Moffat, *Mon. Not. R. Astron. Soc.* **367**, 527–540 (2006), {arXiv:astro-ph/0507222}.
6. J. R. Brownstein, and J. W. Moffat, *Mon. Not. R. Astron. Soc.* **382 (1)**, 29–47 (2007), {arXiv:astro-ph/0702146}.
7. J. W. Moffat, and V. T. Toth, *ArXiv* **0710.0364 [astro-ph]** (2007).
8. J. W. Moffat, and V. T. Toth, *Mon. Not. R. Astron. Soc.* **397**, 1885–1992 (2009), {arXiv:0805.4774[astro-ph]}.
9. J. W. Moffat, and V. T. Toth, *ArXiv* **1001.1564 [gr-qc]** (2010).
10. E. Komatsu, J. Dunkley, M. R. Nolta, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. Limon, L. Page, D. N. Spergel, M. Halpern, R. S. Hill, A. Kogut, S. S. Meyer, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright, *ArXiv* **0803.0547** (2008).
11. U. Seljak, and M. Zaldarriaga, *Astrophys. J.* **469**, 437–+ (1996), {arXiv:astro-ph/9603033}.
12. M. Doran, *Journal of Cosmology and Astroparticle Physics* **10**, 11–+ (2005), arXiv:astro-ph/0302138.
13. D. J. Eisenstein, and W. Hu, *Astrophys. J.* **496**, 605–+ (1998), {arXiv:astro-ph/9709112}.
14. V. Mukhanov, *Physical Foundations of Cosmology*, Cambridge University Press, 2005.
15. T. Padmanabhan, *Structure formation in the universe*, Cambridge University Press, 1993.
16. M. Tegmark, D. J. Eisenstein, M. A. Strauss, D. H. Weinberg, M. R. Blanton, J. A. Frieman, M. Fukugita, J. E. Gunn, A. J. S. Hamilton, G. R. Knapp, R. C. Nichol, J. P. Ostriker, N. Padmanabhan, W. J. Percival, D. J. Schlegel, D. P. Schneider, R. Scoccimarro, U. Seljak, H.-J. Seo, M. Swanson, A. S. Szalay, M. S. Vogeley, J. Yoo, I. Zehavi, K. Abazajian, S. F. Anderson, J. Annis, N. A. Bahcall, B. Bassett, A. Berlind, J. Brinkmann, T. Budavari, F. Castander, A. Connolly, I. Csabai, M. Doi, D. P. Finkbeiner, B. Gillespie, K. Glazebrook, G. S. Hennessy, D. W. Hogg, Ž. Ivezić, B. Jain, D. Johnston, S. Kent, D. Q. Lamb, B. C. Lee, H. Lin, J. Loveday, R. H. Lupton, J. A. Munn, K. Pan, C. Park, J. Peoples, J. R. Pier, A. Pope, M. Richmond, C. Rockosi, R. Scranton, R. K. Sheth, A. Stebbins, C. Stoughton, I. Szapudi, D. L. Tucker, D. E. V. Berk, B. Yanny, and D. G. York, *Phys. Rev. D* **74**, 123507–+ (2006), {arXiv:astro-ph/0608632}.
17. M. Tegmark, M. R. Blanton, M. A. Strauss, F. Hoyle, D. Schlegel, R. Scoccimarro, M. S. Vogeley, D. H. Weinberg, I. Zehavi, A. Berlind, T. Budavari, A. Connolly, D. J. Eisenstein, D. Finkbeiner, J. A. Frieman, J. E. Gunn, A. J. S. Hamilton, L. Hui, B. Jain, D. Johnston, S. Kent, H. Lin, R. Nakajima, R. C. Nichol, J. P. Ostriker, A. Pope, R. Scranton, U. Seljak, R. K. Sheth, A. Stebbins, A. S. Szalay, I. Szapudi, L. Verde, Y. Xu, J. Annis, N. A. Bahcall, J. Brinkmann, S. Burles, F. J. Castander, I. Csabai, J. Loveday, M. Doi, M. Fukugita, J. R. I. Gott, G. Hennessy, D. W. Hogg, Ž. Ivezić, G. R. Knapp, D. Q. Lamb, B. C. Lee, R. H. Lupton, T. A. McKay, P. Kunszt, J. A. Munn, L. O'Connell, J. Peoples, J. R. Pier, M. Richmond, C. Rockosi, D. P. Schneider, C. Stoughton, D. L. Tucker, D. E. Vanden Berk, B. Yanny, and D. G. York, *Astrophys. J.* **606**, 702–740 (2004), {arXiv:astro-ph/0310725}.
18. S. Cole, W. J. Percival, J. A. Peacock, P. Norberg, C. M. Baugh, C. S. Frenk, I. Baldry, J. Bland-Hawthorn, T. Bridges, R. Cannon, M. Colless, C. Collins, W. Couch, N. J. G. Cross, G. Dalton, V. R. Eke, R. De Propris, S. P. Driver, G. Efstathiou, R. S. Ellis, K. Glazebrook, C. Jackson, A. Jenkins, O. Lahav, I. Lewis, S. Lumsden, S. Maddox, D. Madgwick, B. A. Peterson, W. Sutherland, and

K. Taylor, *Mon. Not. R. Astron. Soc.* **362**, 505–534 (2005), {arXiv:astro-ph/0501174}.
19. F. Hoyle, C. M. Baugh, T. Shanks, and A. Ratcliffe, *Mon. Not. R. Astron. Soc.* **309**, 659–671 (1999), {arXiv:astro-ph/9812137}.
20. C. Park, M. S. Vogeley, M. J. Geller, and J. P. Huchra, *Astrophys. J.* **431**, 569–585 (1994).
21. X.-M. Deng, Y. Xie, and T.-Y. Huang, *Phys. Rev. D* **79**, 044014–+ (2009), ArXiv:0901.3730[gr-qc].
22. A. G. Riess, L.-G. Strolger, J. Tonry, S. Casertano, H. C. Ferguson, B. Mobasher, P. Challis, A. V. Filippenko, S. Jha, W. Li, R. Chornock, R. P. Kirshner, B. Leibundgut, M. Dickinson, M. Livio, M. Giavalisco, C. C. Steidel, T. Benítez, and Z. Tsvetanov, *Astrophys. J.* **607**, 665–687 (2004), arXiv:astro-ph/0402512.
23. J. W. Moffat, and V. T. Toth, *in preparation* (2011).

Non-spherical voids: the best alternative to dark energy?

Roberto A Sussman
ICN-UNAM, Circuito Exterior, CU, México DF 04510

September 4, 2011

Abstract

If our cosmic location lies within a large-scale under–dense region or "void", then current cosmological observations can be explained without resorting to a cosmological constant or to an exotic and elusive source like "dark energy". If we further assume this void region to be spherical (as almost all current void models do), then fitting observational data severely constrains our position to be very near the void center, which is a very special and unlikely observation point. We argue in this article that existing spherical void models must be regarded as gross approximations that arise by smoothing out more realistic non–spherical configurations that may fit observations without the limitations imposed by spherical symmetry. In particular, the class of quasi–spherical Szekeres models provides sufficient degrees of freedom to describe the evolution of non–spherical inhomogeneities, including a configuration consisting of several elongated supercluster–like overdense filaments with large underdense regions between them. We summarize a recently published example of such configuration, showing that it yields a reasonable coarse-grained description of realistic observed structures. While the density distribution is not spherically symmetric, its proper volume average yields a spherical density void profile of 250 Mpc that roughly agrees with observations. Also, once we consider our location to lie within a non-spherical void, the definition of a "center" location becomes more nuanced, and thus the constraints placed by the fitting of observations on our position with respect to this location become less restrictive.

1 Introduction.

Inhomogeneous cosmological models have become a valuable tool to analyze cosmological observations without introducing an elusive dark energy source (a comprehensive review on this is found in [1]). The currently preferred inhomogeneous configurations are Gpc-scale under–densities ("voids") based on the spherically symmetric Lemaître-Tolman (LT) models [2, 3], under the assumption that we live close to a center of a cosmic density depression of radius around $1-3$ Gpc [4, 5, 6]. Criticism has been voiced on these void models on the grounds that they violate the Copernican principle, since compliance with the cosmic microwave background (CMB) constraints allows for only one such Gpc structure and the observer location cannot be further away from the origin than ~ 50 Mpc [7] (see also [6]). However, as suggested by more recent work [8, 9], a void of radius 250 Mpc may be sufficient to explain the supernova observations, the power spectrum of the CMB and is also consistent with Big Bang Nucleosynthesis, or Baryon Acoustic Oscillations. By considering void structures of this size the Copernican Principle is not violated, as our Universe may consist of many such structures (the upper size to violate CMB constrains is 300 Mpc [10, 11]). Evidently,

restricting our position to be within 50 Mpc from the center origin of a 250 Mpc void is a less stringent limitation. Notice that these voids are not the smaller voids (30-50 Mpc) seen in the filamentary structure of our Local Universe that roughly correspond to numerical simulations, but would form a structure a larger voids containing the smaller ones yet to be detected by observations.

In a recent article [12] we examined the possibility of using non–spherical void models to describe cosmic inhomogeneities. For this purpose, we considered the class of non–spherical Szekeres solutions of Einstein's equations [13, 14, 15, 16]. By fixing the free parameters of these solutions by means of a thin-shell approximation [10, 11, 17, 18], we obtained a specific model that yields a reasonable coarse-grained description of realistic cosmic structures. Since we define initial conditions at the last scattering surfaces, this model evolves from small early universe initial fluctuations and is consistent with current structure formation scenarios. The model presented in [12] yields an averaged spherically symmetric density distribution with a radial void profile qualitatively analogous to the spherical void models (as those of [8]), hence suggesting that the latter models may be approximate configurations that should emerge after coarse-graining and averaging of under–dense regions of a realistic lumpy non–spherical Universe. Also, the lack of spherical symmetry in the Szekeres model removes the unique invariant nature of the center location of models with this symmetry. Since our being sufficiently near this center is a strong constraint that the fitting of observations place on spherical LT models, this constraint becomes much less restrictive in a non-spherical Szekeres model.

2 Setting up the Szekeres model.

The metric of Szekeres models takes the following form [13]

$$\mathrm{d}s^2 = \mathrm{d}t^2 - \frac{(\Phi' - \Phi\mathcal{E}'/\mathcal{E})^2}{\epsilon - k}\mathrm{d}r^2 - \frac{\Phi^2}{\mathcal{E}}(\mathrm{d}x^2 + \mathrm{d}y^2), \tag{1}$$

where $\Phi = \Phi(t, r)$ and $\Phi' = \partial\Phi/\partial r$, with:

$$\mathcal{E} = \frac{S}{2}\left[\left(\frac{x - P}{S}\right)^2 + \left(\frac{y - Q}{S}\right)^2 + \epsilon\right], \tag{2}$$

while $k(r), S(r), P(r), Q(r)$ are arbitrary functions; ϵ is a constant: the values $\epsilon = 1, 0, -1$ are respectively known as the quasi–spherical, quasi–plane and quasi-hyperbolic Szekeres models (for a detailed discussion on these models see [14, 15, 16]). We consider only the quasispherical case, in which the surfaces marked by r and t constant can be mapped to 2–spheres by a stereographic projection.

Einstein's equations for a dust source associated with (1)–(2) reduce to

$$\dot{\Phi}^2 = -k(r) + \frac{2M(r)}{\Phi}, \tag{3}$$

$$8\pi G\rho = \frac{2M' - 6M\mathcal{E}'/\mathcal{E}}{\Phi^2(\Phi' - \Phi\mathcal{E}'/\mathcal{E})}, \tag{4}$$

where $M(r)$ is an arbitrary function and we assume that $\Phi' \neq \Phi\mathcal{E}'/\mathcal{E}$ holds whenever $M' \neq 3M\mathcal{E}'/\mathcal{E}$,

in order to avoid a shell crossing singularity[16, 26]. The solution of (3) is given by the quadrature

$$\int_0^{\Phi} \frac{\mathrm{d}\widetilde{\Phi}}{\sqrt{-k+2M/\widetilde{\Phi}}} = t - t_B(r). \tag{5}$$

where $t_B(r)$ marks the locus of the big bang (which is, in general, non–simultaneous). We remark that this model has no isometries (it does not admit Killing vectors), but by specializing the free functions we obtain axially and spherically symmetric models as particular cases.

By choosing the r coordinate such that $\bar{r} = \Phi(t_i, r)$, where $t = t_i$ marks the last scattering surface (and dropping the bar to simplify notation), we can eliminates one of the six independent functions of r appearing above. Thus, in order to achieve with a Szekeres model the most realistic possible description of cosmic structures and structure formation, we must prescribe five free functions as initial conditions to specify a unique model. In particular, we will specify the functions S, P, Q, t_B and M. The algorithm that we use in the calculations can be defined as follows:

1. The chosen asymptotic cosmic background is an open Friedman model[1], i.e. $\Omega_m = 0.3$ and $\Lambda = 0$. The background density is then given by

$$\rho_b = \Omega_m \times \rho_{cr} = 0.3 \times \frac{3H_0^2}{8\pi G}(1+z)^3, \tag{6}$$

   where the Hubble constant is $H_0 = 70$ km s^{-1} Mpc^{-1}.

2. We choose $t_B = 0$, hence the age of the Universe (given by (5)) is everywhere the same (as in the homogeneous background Friedmann model) and is equal to $t_i = 471,509.5$ years (see [20] for details).

3. The function $M(r)$ is given by

$$M(r) = 4\pi \frac{G}{c^2} \int_0^r \rho_b (1 + \delta\bar{\rho})\, \bar{r}^2 \, \mathrm{d}\bar{r},$$

   where $\delta\bar{\rho} = -0.005\mathrm{e}^{-(\ell/100)^2} + 0.0008\mathrm{e}^{-[(\ell-50)/35]^2} + 0.0005\mathrm{e}^{-[(\ell-115)/60]^2} + 0.0002\mathrm{e}^{-[(\ell-140)/55]^2}$, and $\ell \equiv r/$ 1 kpc.

4. The function $k(r)$ can be calculated from (5).

5. The functions Q, P, and S are prescribed in order to provide the best possible coarse–grained description of the density distribution of our observed local Cosmography by means of a thin shell approximation (see [12]).

6. Once the model is specified, its evolution is calculated from eq. (3) and the density distribution at the current instant is evaluated from (4).

[1] Asymptotic spatial flatness is no longer required if homogeneity is relaxed [6, 19].

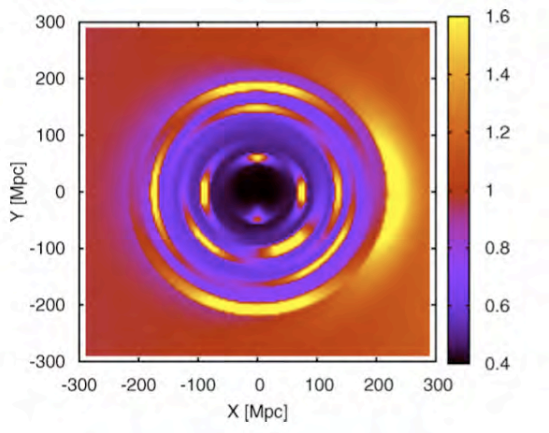


Figure 1: The present-day color-coded density distribution ρ/ρ_0 (where ρ_0 is density of the homogeneous background model). Brighter colors indicate a high-density region, darker low-density region3.

3 How realistic this model can be?

The density distribution for our model (depicted in Fig. 1 in intuitive Cartesian coordinates [21, 22]) follows from our choice of the functions $\{M, t_B, Q, P, S\}$. If new data would arise showing a different density pattern, we can always adjust it appropriately by selecting different functions that would change the position, size, and the amplitude of the overdensities (see [21, 22] for a detailed discussion).

As shown in Fig. 1, the model under consideration contains structures such as voids and elongated supercluster–like overdensities. It has large overdensities around $\sim$ 200 Mpc (towards the left of the figure) that compensate the underdense regions and allow the model to be practically homogeneous at $r > 300$ Mpc. Actual observations reveal very massive matter concentrations – the Shapley Concentration roughly at the distance of 200 Mpc, or the Great Sloan Wall at the distance of 250-300 Mpc. In the opposite direction on the sky we find the Pisces–Cetus and Horologium–Reticulum, which are massive matter concentrations located at a similar distance. We refer the reader to Fig. 44 of Ref. [24], which provides a density map of the Local Universe reconstructed from the 2dF Galaxy Redshift Survey Survey using Delaunay Tessellation Field Estimator[2]. Also, the inner void seen in Fig. 1 is consistent with what is observed in the Local Universe – it appears that our Local Group is not located in a very dense region of the Universe, rather it is located in a less dense region surrounded by large overdensities like the Great Attractor on one side and the Perseus–Piscis supercluster on the other side. Both are located at around 50 Mpc — see Fig. 19 of [25] that provides the density reconstruction of the Local Universe using the POTENT analysis.

While still far from a perfect "realistic" description, the density pattern displayed in Fig. 1 exhibits the main features of our local Universe. It should be therefore treated as a "coarse-grained" approximation to study local cosmic dynamics by means of a suitable exact solution of Einstein's equations. Such approximation is, evidently, far less idealized than the gross one that follows from

[2]This figure is also available at http://en.wikipedia.org/wiki/File:2dfdtfe.gif

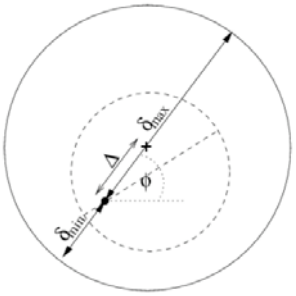


Figure 2: Schematic representation of locations that can be considered "centers" in a quasi–spherical Szekeres model: the local isotropic observer at the origin $r = 0$ (denoted by a black dot) where shear vanishes and the geometric center of the larger sphere depicted by a cross. The distance between these locations is denoted by Δ.

spherically symmetric LT models.

4 Position of the "center"

As a consequence of the lack of spherical symmetry, the model under consideration lacks an invariant and unique characterization of a center worldline. Instead, for every 2–sphere corresponding to a fixed value of r at an instant $t =$ constant, we have (at least) two locations that can be considered appropriate generalizations of the spherically symmetric center: the worldline marked by the coordinate "origin" $r = 0$ where the shear tensor vanishes, which defines a locally isotropic observer (cf. eq (16.29) of Ref. [26]), and the "geometric" center of the 2–sphere whose surface area is $4\pi\Phi^2$.

As shown in Fig. 2, the fact that the 2–spheres of constant r in a quasi-spherical Szekeres model are non-concentric implies that the geometric center of these spheres and $r = 0$ do not coincide. As a consequence, the distance from this origin to the surface of the sphere depends on the direction marked by the angles (θ, ϕ) of the stereographic projection (see equation (3) of Ref. [12]):

$$\delta(r, \theta, \phi) = \int_0^r \mathrm{d}\tilde{r} \frac{\Phi' - \Phi\mathcal{E}'/\mathcal{E}}{\sqrt{1-k}}. \tag{7}$$

Hence, the displacement Δ between the origin and the geometric center of a sphere of radius r is

$$\Delta = \frac{\delta_{\max} - \delta_{\min}}{2},$$

where $\delta_{\max} = \max(\delta)$, $\delta_{\min} = \min(\delta)$. As can be seen from equation (3) of [12], the maximal and minimal value of $\mathcal{E}'/\mathcal{E}$ for our model (where $S' = 0$) corresponds to $\theta = \pi/2$. The distance, δ, as a function of ϕ for voids of various radii is depicted by figure 3 of [12], showing that a sphere whose present-day area radius is $\Phi = 100$ Mpc the model under consideration yields a displacement of $\Delta = 36$ Mpc towards $\phi \approx 80^\circ$ direction. While for $\Phi = 250$ Mpc we have $\Delta = 62$ Mpc towards $\varphi \approx 120^\circ$. Fitting observations in spherically symmetric models restricts our cosmic location to be within a given maximal separation from a location that is both, the geometric center of the void and

the locally isotropic observer ($\Delta = 0$). It is reasonable to expect that similar distance restrictions with respect to the local isotropic observer should emerge in fitting observations with a Szekeres model, but in the latter models this observer is not the only center and may be far away from the geometric center of the void, and thus our location would be less special and improbable than in spherically symmetric models where both locations coincide.

5 Averaging

As shown in Ref. [23], the proper 3–dimensional volume in space slices orthogonal to the 4–velocity (t = constant) in a Szekeres model is

$$V_{\mathcal{D}} = \int_0^{r_{\mathcal{D}}} \mathrm{d}r \int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{d}y \sqrt{-g} = 4\pi \int_0^{r_{\mathcal{D}}} \mathrm{d}r \frac{\Phi^2 \Phi'}{\sqrt{1-k}} \equiv 4\pi R_{\mathcal{D}}, \tag{8}$$

and thus, the proper volume averaged density is spherically symmetric (*i.e.* independent of x and y), even if the density itself is far from a spherical distribution:

$$\langle \rho \rangle (r_{\mathcal{D}}) = \frac{1}{V_{\mathcal{D}}} \int_0^{r_{\mathcal{D}}} \mathrm{d}r \int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \rho \mathrm{d}y \sqrt{-g} = \frac{1}{8\pi G R_{\mathcal{D}}} \int_0^{r_{\mathcal{D}}} \mathrm{d}r \frac{2M'}{\sqrt{1-k}}. \tag{9}$$

The radial profile of this spherical volume–averaged density distribution evaluated as a function of $r_{\mathcal{D}}$, is displayed by Fig. 3. The spherical symmetry of the averaged density distribution implies that the the averaging process has smoothed out the "angular" (*i.e.* x, y) dependence of a highly non–spherical coarse grained density distribution. Since the resulting averaged distribution $\langle \rho \rangle (r_{\mathcal{D}})$ is equivalent to a spherical cosmic void whose radius is approximately 250 Mpc (as in Ref. [8]), the latter type of void models can be thought of as rough averages of more realistic non–spherical configurations. As a consequence, the use of a Szekeres model seems to suggest that results obtained by means of spherical LT models may be robust: while local non–spherical information could still provide important refinements, and is needed for computations involving null geodesics (specially when fitting CMB constraints), it is likely that basic bottom line information is already contained in the spherical voids constructed with LT models.

6 Conclusions

The model we have presented is one among the first attempts in using the Szekeres solution as a theoretical and empiric tool to study and interpret cosmological observations [28, 29, 9, 30]. This opens new possibilities for inhomogeneous cosmologies, as this is the most general available cosmological exact inhomogeneous and anisotropic solution of Einstein's equations. The model provides a more nuanced and much less restrictive description of the need to constrain our location with respect to a center location. It is also a concrete example that illustrates the possibility that a mildly increasing void profile (required by observations) can emerge if local structures are coarse-grained and then averaged. Of course, notwithstanding these appealing features, the model and its assumptions must be subjected to hard testing by data from the galaxy redshift surveys, and evidently the more comprehensive this data can be the better it can be used for this purpose.

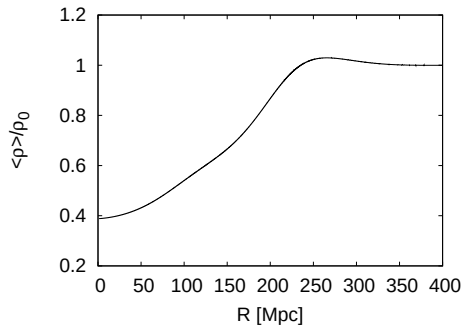


Figure 3: Radial profile of the spherically symmetric averaged distribution (normalized by the background density ρ_0)

Unfortunately current surveys like 2dF of SDSS do not cover the whole sky and only focus on small angular regions of it. However in the near future this limitation may be overcome – for example, Sky Mapper[3] aims to cover the whole southern sky which will provide sufficient data to test possibilities suggested and elaborated in this work. A more comprehensive and detailed article on the model proposed here is currently under elaboration and will be submitted soon for publication.

References

[1] K. Bolejko *et al.*, *Structures in the Universe by Exact Methods – Formation, Evolution, Interactions.*, Cambridge University Press, Cambridge, 2009

[2] G. Lemaître, Ann. Soc. Sci. Bruxelles A53 (1933) 51; English translation, with historical comments: Gen. Rel. Grav. 29 (1997) 637.

[3] R.C. Tolman, Proc. Nat. Acad. Sci. USA 20 (1934) 169; reprinted, with historical comments: Gen. Rel. Grav. 29 (1997) 931.

[4] H. Alnes, M. Amarzguioui, ØGrøn, Phys. Rev. D73 (2006) 08351.

[5] J. García-Bellido, T. Haugbølle, J. Cosmol. Astropart. Phys. 04 (2008) 003.

[6] K. Bolejko and J. S. B. Wyithe, J. Cosmol. Astropart. Phys. 02 (2009) 020.

[7] H. Alnes, M. Amarzguioui, Phys. Rev. D75 (2007) 023506.

[8] S. Alexander et al., J. Cosmol. Astropart. Phys. 09, 025 (2009).

[9] K. Bolejko and M.-N. Célérier, Phys. Rev. D82 (2010) 103510

[10] K.T. Inoue and J. Silk, Astrophys. J. 648 (2006) 23.

[3] http://msowww.anu.edu.au/skymapper/

[11] K.T. Inoue and J. Silk, Astrophys. J. 664 (2007) 650.

[12] K. Bolejko and R.A. Sussman, Phys Lett B 697 265–270.

[13] P. Szekeres, Commun. Math. Phys. 41 (1975) 55.

[14] C. Hellaby, A. Krasiński, Phys. Rev. D77 (2008) 023529

[15] A. Krasiński, Phys. Rev. D78 (2008) 064038.

[16] C. Hellaby, A. Krasiński, Phys. Rev. D66 (2002) 084011

[17] K.L. Thompson and E.T. Vishniac, Astrophys. J. 313 (1987) 517.

[18] K. Tomita, Astrophys. J. 529 (2000) 38.

[19] C. Clarkson, M. Regis, arXiv:1007.3443 (2010)

[20] P.J.E. Peebles, The Large-Scale Structure of the Universe. Princton University Press, Princton (1980)

[21] K. Bolejko, Phys. Rev. D73 (2006) 123508.

[22] K. Bolejko, Phys. Rev. D75 (2007) 043508.

[23] K. Bolejko, Gen. Rel. Grav. 41 (2009) 1585.

[24] R. van de Weygaert and W. Schaap in Data Analysis in Cosmology, ed. V. Martínez, E. Saar, E. Martínez-González, M. Pons-Bordería, Springer-Verlag, Berlin, Lecture Notes in Physics 665 (2009) p. 291.

[25] A. Dekel et al., Astrophys. J. 522 (1999) 1.

[26] J. Plebański, A. Krasiński, An introduction to general relativity and cosmology. Cambridge University Press, Cambridge (2006).

[27] T. Buchert, Gen. Rel. Grav. 40 (2008) 467; R. Zalaletdinov, Int. J. Mod. Phys. A23 (2008) 1173

[28] M. Ishak, J. Richardson, D. Garred, D. Whittington, A. Nwankwo, R. Sussman, Phys. Rev. D78 (2008) 123531.

[29] A. Nwankwo, J. Thompson, M. Ishak, arXiv1005.2989 (2010).

[30] A. Krasiński, K. Bolejko, arXiv:1007.2083 (2010).

Lecture International Conference 2010
se terminó de editar en agosto de 2012
ebook

The International Conference on Two Cosmological Models was held at the Universidad Iberoamericana in Mexico City, Mexico, from November 17th to 19th, 2010, as a forum devoted to the study and discussion of two important problems of modern cosmology. More than 38 speakers from Brazil, Canada, Chile, France, Germany, India, Mexico, the Netherlands, New Zealand, and the United States participated in an open discussion on the following two topics:

1) The concept of dark matter as a possible explanation of the rotation velocity of galaxies and galaxy clusters in the context of Newtonian dynamics; and the alternative explanation through Einstein's general relativity, without dark matter.

2) The concept of dark energy as a possible explanation of the apparent acceleration of the expansion of the universe; and the alternative explanation through Einstein's gravitational theory, without dark energy. The overall impact of the event was more than satisfactory. These Proceedings contain the lectures on the topics covered in the International Conference on Two Cosmological Models, except for two of them, who could not send us the written version of their lecture.