Space-Time-Matter

Modern Kaluza-Klein Theory

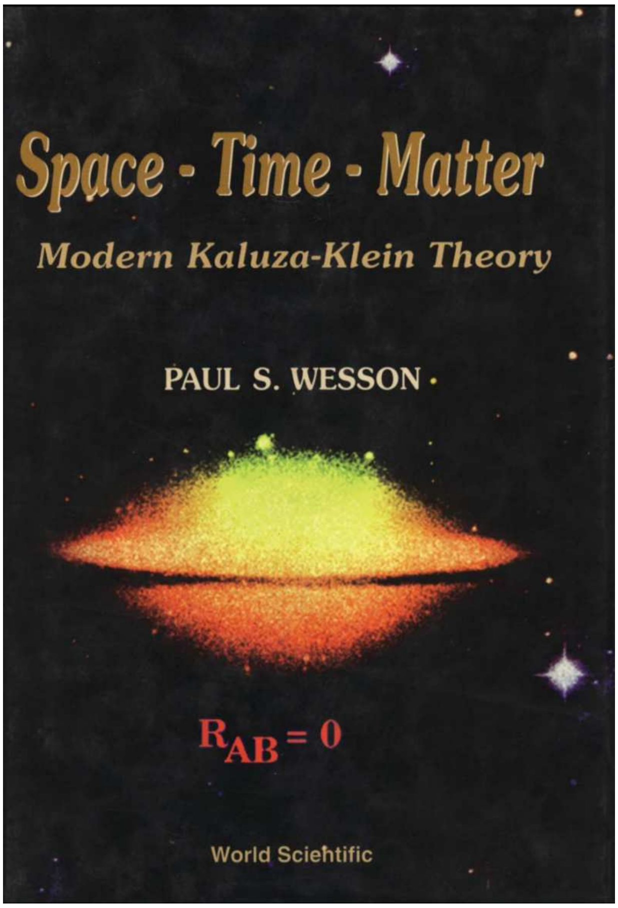
Paul S. Wesson

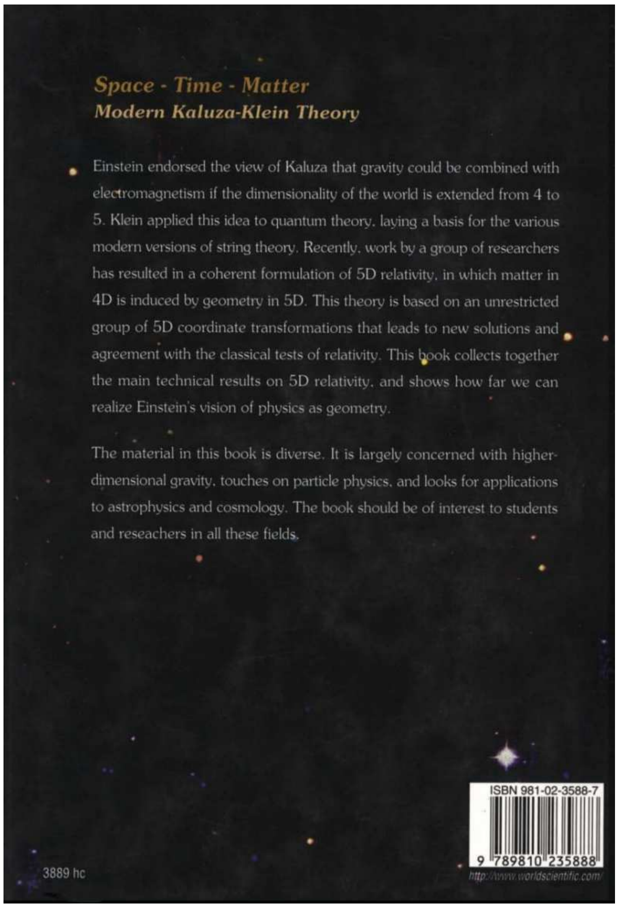
Contents

Chapter 0

Preliminary

0.1 Capa





Space - Time - Matter
Modern Kaluza-Klein Theory

Paul S. Wesson

University of Waterloo, Ontario, Canada and
Hansen Physics Labs, Stanford University

World Scientific

Singapore • New Jersey • London • Hong Kong

Published by

World Scientific Publishing Co. Re. Ltd.

P 0 Box 128, Farrer Road, Singapore 912805

USA *office*: Suite 1B, 1060 Main Street, River Edge, NJ 07661

UK *office*: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

SPACE, TIME, MATTER: MODERN KALUZA-KLEIN THEORY

ISBN 981-02-3588-7

Printed in Singapore.

0.2 Preface

Einstein endorsed the view of Kaluza, that gravity could be combined with electromagnetism if the dimensionality of the world is extended from 4 to 5. Klein applied this idea to quantum theory, laying a basis for the various modern versions of string theory. Recently, work by a group of researchers has resulted in a coherent formulation of 5D relativity, in which matter in 4D is induced by geometry in 5D. This theory is based on an unrestricted group of 5D coordinate transformations that leads to new solutions and agreement with the classical tests of relativity. This book collects together the main technical results on 5D relativity, and shows how far we can realize Einstein's vision of physics as geometry.

Space, time and matter are physical concepts, with a long but somewhat subjective history. Tensor calculus and differential geometry are highly developed mathematical formalisms. Any theory which joins physics and algebra is perforce open to discussions about interpretation, and the one presented in this book leads to new issues concerning the nature of matter. The present theory should not strictly speaking be called Kaluza-Klein: KK theory relies on conditions of cylindricity and compactification which are now removed. The theory should also, while close to it in some ways, not be confused with general relativity: GR theory has an explicit energy-momentum tensor for matter while now there is none. What we call matter in 4D spacetime is the manifestation of the fifth dimension, hence the phrase induced-matter theory sometimes used in the literature. However, there is nothing sacrosanct about 5D. The field equations take the same form in ND, and N is to be chosen with a view to physics. Thus, superstrings (10D) and supergravity (11D) are valid constructs. However, practical physical applications are expected to be forthcoming only if there is physical understanding of the nature of the extra dimensions and the extra coordinates. In this regard, space-time-matter theory is uniquely fortunate. This because (unrestricted) 5D Riemannian geometry turns out to be just algebraically rich enough to unify gravity and electromagnetism with their sources of mass and charge. In other words, it is a Machian theory of mechanics.

There is now a large and rapidly growing literature on this theory, and the author is aware that what follows is more like a textbook on basics than a review of recent discoveries. It should also be stated that much of what follows is the result of a group effort over time. Thus credit is due especially to H. Liu, B. Mashhoon and J. Ponce de Leon for their solid theoretical work; to C.W.F. Everitt who sagely kept us in contact with experiment; and to A. Billyard, D. Kalligas, J.M. Overduin and W. Sajko, who as graduate students cheerfully tackled problems that would have made their older colleagues blink. Thanks also go to S. Chatterjee, A. Coley, T. Fukui and R. Tavakol for valuable contributions. However, the responsibility for any errors or omissions rests with the author.

The material in this book is diverse. It is largely concerned with higher-dimensional gravity, touches particle physics, and looks for application to astrophysics and cosmology. Depending on their speciality, some workers may not wish to read this book from cover to cover. Therefore the material has been arranged in approximately self-contained chapters, with a bibliography at the end of each. The material does, of course, owe its foundation to Einstein. However, it will be apparent to many readers that it also owes much to the ideas of his contemporary, Eddington.

Paul S. Wesson

0.3 Contents

Chapter 1

Concepts and Theories of Physics

"Physics should be beautiful"

(Sir Fred Hoyle, Venice, 1974)

1.1 Introduction

Physics is a logical activity, which unlike some other intellectual pursuits frowns on radical departures, progressing by the introduction of elegant ideas which give a better basis for what we already know while leading to new results. However, this inevitably means that the subject at a fundamental level is in a constant state of reinterpretation. Also, it is often not easy to see how old concepts fit into a new framework. A prime example is the concept of mass, which has traditionally been regarded as the source of the gravitational field. Historically, a source and its field have been viewed as separate things. But as recognized by a number of workers through time, this distinction is artificial and leads to significant technical problems. Our most successful theory of gravity is general relativity, which traditionally has been formulated in terms of a set of field equations whose left-hand side is geometrical (the Einstein tensor) and whose right-hand side is material (the energy-momentum tensor). However, Einstein himself realized soon after the formulation of general relativity that this split has drawbacks, and for many years looked for a way to transpose the "base-wood" of the right-hand side of his equations into the "marble" of the left-hand side. Building on ideas of Kaluza and Klein, it has recently become feasible to realize Einstein's dream, and the present volume is mainly a collection of technical results, which shows how this can be done. The basic idea is to unify the source and its field using the rich algebra of higher-dimensional Riemannian geometry. In other words: space, time and matter become parts of geometry.

This is an idea many workers would espouse, but to be something more than an academic jaunt we have to recall the two conditions noted above. Namely, we have to recover what we already know (with an unavoidable need for reinterpretation); and we have to derive something new with at least a prospect of testability. The present chapter is concerned with the first of these, and the succeeding chapters mainly with the second. Thus the present chapter is primarily a review of gravitation and particle physics as we presently understand these subjects. Since this is mainly known material, these accounts will be kept brief, and indeed those readers who are familiar with these subjects may wish to boost through them. However, there is a theme in the present chapter, which transcends the division of physics into theories of macroscopic and microscopic scope. This is the nature and origin of the so-called fundamental constants. These are commonly taken as indicators of what kind of theory is under consideration (e.g., Newton's constant is commonly regarded as typical of classical theory and Planck's constant as typical of quantum theory). But at least one fundamental constant, the speed of light, runs through all modern physical theories; and we cannot expect to reach a meaningful unification of the latter without a proper understanding of where the fundamental constants originate. In fact, the chapters after this one make use of a view that it is necessary to establish and may be unfamiliar to some workers: the fundamental constants are not really fundamental, their main purpose being to enable us to dimensionally transpose certain material quantities so that we can write down consistent laws of physics.

1.2 Fundamental constants

A lot has been written on these, and there is a large literature on unsuccessful searches for their possible variations in time and space. We will be mainly concerned with their origin and status, on which several reviews are available. Notably there are the books by Wesson (1978), Petley (1985) and Barrow and Tipler (1986); the conference proceedings edited by McCrea and Rees (1983); and the articles by Barrow (1981) and Wesson (1992). We will presume a working physicist's knowledge of the constants concerned, and the present section is to provide a basis for the discussions of physical theory which follow.

The so-called fundamental constants are widely regarded as a kind of distillation of physics. Their dimensions are related to the forms of physical laws, whose structure can in many cases be recovered from the constants by dimensional analysis. Their sizes for some choice of units allow the physical laws to be evaluated and compared to observation. Despite their perceived fundamental nature, however, there is no theory of the constants as such. For example, there is no generally accepted formalism that tells us how the constants originate, how they relate to one another, or how many of them are necessary to describe physics. This lack of background seems odd for parameters that are widely regarded as basic.

The constants we will be primarily concerned with are those that figure in gravity and particle physics. It is convenient to collect the main ones here, along with their dimensions and approximate sizes in c.g.s. units:

Speed of light	c	$L\,T^{-1}$	3.0×10^{10}
Gravitational constant	G	$M^{-1}\,L^3\,T^{-2}$	6.7×10^{-8}
Planck's constant	h	$M\,L^2\,T^{-1}$	6.6×10^{-27}
Electron charge (modulus)	e	$M^{1/2}\,L^{3/2}\,T^{-1}$	4.8×10^{-10}

Here e is measured in electrostatic or Gaussian units. We will use e.s.u. in the bulk of what follows, though S.I. will be found useful in places. The two systems of units are of course related by $4\pi\varepsilon_0$, where the permittivity of free space is $\varepsilon_0 = 8.9\times10^{-12}C^2m^{-3}s^2Kg^{-1}$. In S.I. $e = 1.6\times10^{-19}C$ (Coulombs: see Jackson 1975, pp. 29, 817; and Griffiths 1987, p. 9). The permeability of free space μ_0, is not an independent constant because $c^2 \equiv 1/\varepsilon_0\mu_0$. The above table suggests that we need to understand 3 overlapping things: constants, dimensions and units.

One common view of the constants is that they define asymptotic states. Thus c is the maximum velocity of a massive particle moving in flat spacetime; G defines the limiting potential for a mass that does not form a black hole in curved spacetime; ε_0 is the empty-space or vacuum limit of the dielectric constant; and h defines a minimum amount of energy (alternatively $\hbar \equiv h/2\pi$ e defines a minimum amount of angular momentum). This view is acceptable, but somewhat begs the question of the constants' origin.

Another view is that the constants are necessary inventions. Thus if a photon moves away from an origin and attains distance r in time t, it is necessary to write $r = ct$ as a way of reconciling the different natures of space and time. Or, if a test particle of mass m_1 moves under the gravitational attraction of another mass m_2 and its acceleration is d^2r/dt^2 at separation r, it is observed that

$m_1 d^2r/dt^2$ is proportional to $m_1 m_2/r^2$, and to get an equation out of this it is necessary to write $d^2r/dt^2 = Gm_2/r^2$ as a way of reconciling the different natures of mass, space and time. A similar argument applies to the motion of charged bodies and ε_0. In quantum theory, the energy E of a photon is directly related to its frequency ν, so we necessarily have to write $E = h\nu$. The point is, that given a law of physics which relates quantities of different dimensional types, constants with the dimensions $c = LT^{-1}$, $G = M^{-1}L^3T^{-2}$, $\varepsilon_0 = Q^2M^{-1}L^{-3}T^2$ and $h = ML^2T^{-1}$ are obligatory.

This view of the constants is logical, but disturbing to many because it means they are not really fundamental and in fact largely subjective in origin. However, it automatically answers the question raised in the early days of dimensional analysis as to why the equations of physics are dimensionally homogeneous (e.g. Bridgman 1922). It also explains why subsequent attempts to formalize the constants using approaches such as group theory have led to nothing new physically (e.g. Bunge 1971). There have also been notable adherents of the view that the fundamental constants are not what they appear to be. Eddington (1929, 1935, 1939) put forward the opinion that while an external world exists, our laws are subjective in the sense that they an constructed to match our own physical and mental modes of perception. Though he was severely criticized for this opinion by physicists and philosophers alike, recent advances in particle physics and relativity make it more palatable now than before. Jeffreys (1973, pp. 87-94, 97) did not see eye to eye with Eddington about the sizes of the fundamental constants, but did regard some of them as disposable. In particular, he pointed out that in electrodynamics c merely measures the ratio of electrostatic and electromagnetic units for charge. Hoyle and Narlikar (1974, pp. 97, 98) argued that the c^2 in the common relativistic expression $(c^2t^2 - x^2 - y^2 - z^2)$ should not be there, because “there is no more logical reason for using a different time unit than there would be for measuring x, y, z in different units”. They stated that the velocity of light is unity, and its size in other units is equivalent to the definition $1\ s = 299\ 792\ 500\ m$, where the latter number is manmade. McCrea (1986, p. 150) promulgated an opinion that is exactly in line with what was discussed above, notably in regard to c, h and G, which he regarded as “conversion constants and nothing more”. These comments show that there is a case that can be made for removing the fundamental constants from physics.

Absorbing constants in the equations of physics has become commonplace in recent years, particularly in relativity where the algebra is usually so heavy that it is undesirable to encumber it with unnecessary symbols. Formally, the rules for carrying this out in a consistent fashion are well known (see e.g. Desloge 1984). Notably, if there are N constants with N bases, and the determinant of the exponents of the constants' dimensions is nonzero so they are independent, then their magnitudes can be set to unity. For the constants c, G, ε_0, h with bases M, L, T, Q it is obvious that ε_0 and Q can be removed this way. (Setting $\varepsilon_0 = 1$ gives Heaviside-Lorentz units, which are not the same as setting $4\pi\varepsilon_0 = 1$ for Gaussian units, but the principle is clearly the same: see Griffiths, 1987, p. 9.) The determinant of the remaining dimensional combinations $M^0L^1T^{-1}$, $M^{-1}L^3T^{-2}$, $M^1L^2T^{-1}$ is finite, so the other constants c, G, h can be set to unity. Conceptually, the absorbing of constants in this way prompts 3 comments. (a) There is an overlap and ambiguity between the idea of a base dimension and the idea of a unit. All of mechanics can be expressed with dimensional bases M, L, T; and we have argued above that these originate because of our perceptions of mass, length and time as being different things. We could replace one or more of these by another base (e.g. in engineering force is sometimes used as a base), but there will still be 3. If we extend mechanics to include electrodynamics, we need to add a new base Q. But the principle is clear, namely that the base dimensions reflect the nature and extent of physical theory. In contrast, the idea of a unit is less conceptual but more practical. We will discuss units in more detail below, but for now we point up the distinction by noting that a constant can have different sizes depending on the choice of units while retaining the same dimensions. (b) The process of absorbing constants cannot be carried arbitrarily far. For example, we cannot set $e = 1$, $\hbar = 1$ and $c = 1$ because it makes the electrodynamic fine-structure constant $\alpha \equiv e^2/\hbar c$ equal to 1, whereas in the real world it is observed to be approximately $1/137$. This value actually has to do with the peculiar status of e compared to the other constants (see below), but the caution is well taken. (c) Constants mutate with time. For example, the local acceleration of gravity g was apparently at one time viewed as a 'fundamental' constant, because it is very nearly the same at all places on the Earth's surface. But today we know that $g = GM_E/r_E^2$ in terms of the mass and radius of the Earth, thus redefining g in more basic terms. Another example is that the gravitational coupling constant in general relativity is not really G but the combination $8\pi G/c^4$ (Section 1.3), and more examples are forthcoming from particle physics (Section 1.4). The point of this and the preceding comments is that where the fundamental constants are concerned, formalism is inferior to understanding.

To gain more insight, let us discuss in greater detail the relation between base dimensions and units, concentrating on the latter. There are 7 base dimensions in widespread use (Petley 1985, pp. 26-29). Of these 3 are the familiar M, L, T of mechanics. Then electric current is used in place of Q. And the other 3 are temperature, luminous intensity and amount of substance (mole). As noted above, we can swap dimensional bases if we wish as a matter of convenience, but the status of physics fixes their number. By contrast, choices of units are infinite in number. At present there is a propensity to use the S.I. system (Smith 1983). While not enamoured by workers in astrophysics and certain other disciplines because of the awkwardness of the ensuing numbers, it is in widespread use for laboratory-based physics. The latter requires well-defined and reproducible standards, and it is relevant to review here the status of our basic units of time, length and mass.

The second in S.I. units is defined as 9 192 631 770 periods of a microwave oscillator running under well-defined conditions and tuned to maximize the transition rate between two hyperfine levels in the ground state of atoms of ^{133}Cs moving without collisions in a near vacuum. This is a fairly sophisticated definition, which is used because the caesium clock has a long-term stability of 1 part in 10^{14} and an accuracy of reproducibility of 1 part in 10^{13}. These specifications are better than those of any other apparatus, though in principle a water clock would serve the same purpose. So much for a unit of time. The metre was originally defined as the distance between two scratch marks on a bar of metal kept in Paris. But it was redefined in 1960 to be 1 650 763.73 wavelengths of one of the orange-red lines in the spectrum of a ^{83}Kr lamp running under certain well-defined conditions. This standard, though, was defined before the invention of the laser with its high degree of stability, and is not so good. A better definition of the metre can be made as the distance traveled by light in vacuum in a time of 1/2 997 924.58 (caesium clock) seconds. Thus we see that a unit of length can be defined either autonomously or in conjunction with the speed of light. The kilogram started as a lump of metal in Paris, but unlike its compatriot the metre continued in use in the form of carefully weighed copies. This was because Avogadro's number, which gives the number of atoms in a mass of material equal to the atomic number in grams, was not known by traditional means to very high precision. However, it is possible to obtain a better definition of the kilogram in terms of Avogadro's number derived from the lattice spacing of a pure crystal of a material like ^{26}Si, where the spacing can be determined by X-ray diffraction and optical interference. Thus, a unit of mass can be defined either primitively or in terms of the mass of a crystal of known size. We conclude that most accuracy can be achieved by defining a unit of time, and using this to define a unit of length, and then employing this to obtain a unit of mass. However, more direct definitions can be made for all of these quantities, and there is no reason as far as units are concerned why we should not absorb c, G and h.

This was actually realized by Planck, who noted that their base dimensions are such as to allow us to define 'natural' units of mass, length and time. (See Barrow 1983: similar units were actually suggested by Stoney somewhat earlier; and some workers have preferred to absorb $\hbar$ rather than h.) The correspondence between natural or Planck units and the conventional gram, centimetre and second can be summarized as follows:

$$1m_p \equiv \left(\frac{\hbar c}{G}\right)^{1/2} = 2.2 \times 10^{-5} g \qquad 1g = 4.6 \times 10^4 m_p$$

$$1l_p \equiv \left(\frac{G\hbar}{c^3}\right)^{1/2} = 1.6 \times 10^{-33} cm \qquad 1cm = 6.3 \times 10^{32} l_p$$

$$1t_p \equiv \left(\frac{G\hbar}{c^5}\right)^{1/2} = 5.4 \times 10^{-44} s \qquad 1s = 1.9 \times 10^{43} t_p$$

In Planck units, all of the constants c, G and $\hbar$ become unity and they consequently disappear from the equations of physics.

This is convenient but it involves a choice of *units* only and does not necessarily imply anything more. It has often been stated that a consistent theory of quantum gravity that involves c, G and $\hbar$ would naturally produce particles of the Planck mass noted above. However, this is theoretically unjustified based on what we have discussed; and seems to be practically supported by the observation that the universe is not dominated by 10^{-5} g black holes. A more significant view is that all measurements and observations involve comparing one thing with another thing of similar type to produce what is ultimately a dimensionless number (see Dicke 1962; Bekenetein 1979; Barrow 1981; Smith 1983; Wesson 1992). The latter can have any value, and are the things that physics needs to explain. For example, the electromagnetic fine-structure constant $\alpha \equiv e^2/\hbar c \cong 1/137$ needs to be explained, which is equivalent to saying that the electron charge needs to be explained (Griffiths 1987). The 'gravitational fine-structure constant' $Gm_p^2/\hbar c \cong 5 \times 10^{-39}$ needs to be explained, which is equivalent to saying that the mass of the proton needs to be explained (Carr and Rees 1979). And along the same lines, we need to explain the constant involved in the observed correlation between the spin angular momenta and masses of astronomical objects, which is roughly $GM^2/Jc \cong 1/300$ (Wesson 1983). In other words, we get no more out of dimensional analysis and a choice of units than is already present in the underlying equations, and neither technique is a substitute for proper physics.

The physics of explaining the charge of the electron or the mass of a proton, referred to above, probably lies in the future. However, some comments can be made now. As regards e, it is an observed fact that α is energy or distance-dependent. Equivalently, e is not a fundamental constant in the same class as c, h and G. The current explanation for this involves vacuum polarization, which effectively screens the charge of one particle as experienced by another (see Section 1.4). This mechanism is depressingly mechanical to some field theorists, and in attributing an active role to the vacuum would have been anathema to Einstein. [There are also alternative explanations for it, such as the influence of a scalar field, as discussed in Nodvik (1985) and Chapter 5.] However, the philosophy of trying to understand the electron charge, rather than just accepting it as a given, has undoubted merit. The same applies to the masses of the elementary particles, which however are unquantized and so present more of a challenge. The main question is not whether we wish to explain charges and masses, but rather what is the best approach.

In this regard, we note that both are geometrizable (Hoyle and Narlikar 1974; Wesson 1992). The rest mass of a particle m is the easiest to treat, since using G or h we can convert m to a length:

$$x_m \equiv \frac{Gm}{c^2} \qquad or \qquad x_m \equiv \frac{h}{mc} \quad .$$

Physically, the choice here would conventionally be described as one between gravitational or atomic units, a ploy which has been used in several theories that deal with the nature of mass (see Wesson 1978 for a review). Mathematically, the choice is one of coordinates, provided we absorb the constants and view mass as on the same footing as time and space (see Chapter 7). The electric charge of a particle q is harder to treat, since it can only be geometrized by including the gravitational constant via $x_q \equiv (G/c^4)^{1/2} q$. This, together with the trite but irrefutable fact that masses can carry charges but not the other way round, suggests that mass is more fundamental than charge.

1.3 General relativity

In the original form of this theory due to Einstein, space is regarded as a construct in which only the relations between objects have meaning. The theory agrees with all observations of gravitational phenomena, but the best books that deal with it are those which give a fair treatment of the theory's conceptual implications. Notably, those by Weinberg (1972), Misner, Thome and Wheeler (1973), Rindler (1977) and Will (1993). We should also mention the book by Jammer (1961) on concepts of mass; and the conference proceedings edited by Barbour and Pfister (1995) on the idea due to Mach that mass locally depends on the distribution of matter globally. The latter was of course a major motivation for Einstein, and while not incorporated into standard general relativity is an idea that will reoccur in subsequent chapters.

The theory is built on 10 dimensionless potentials which are the independent elements in a 4 x 4 metric tensor $g_{\alpha\beta}$ $(\alpha, \beta = 0 - 3)$. These define the square of the distance between 2 nearby points in 4D via $ds^2 = g_{\alpha\beta} dx^\alpha dx^\beta$. (Here a repeated index upstairs and downstairs is summed over, and below we will use the metric tensor to raise and lower indices on other tensors.) The coordinates x^α are in a local limit identified as $x^0 = ct$, $x^1 = x$, $x^2 = y$, $x^3 = z$ using Cartesians. However, because the theory employs tensors and therefore gives relations valid in any system of coordinates (covariance), the space and time labels may be mixed up and combined arbitrarily. Thus space and time are not distinct entities. Also, the role of the speed of light c is to dimensionally transpose a quantity with the dimension T to one with dimension L, so that all 4 of x^α may be treated on the same footing. Partial derivatives with respect to the x^α can be combined to produce the Christoffel symbol $\Gamma^\gamma_{\alpha\beta}$, which enables one to create a covariant derivative such that the derivative of a vector is now given by $\nabla_\alpha V_\beta = \partial V_\alpha / \partial x^\beta - \Gamma^\gamma_{\alpha\beta} V_\gamma$. From $g_{\alpha\beta}$, and its derivatives, one can obtain the Ricci tensor $R_{\alpha\beta}$, the Ricci scalar R and the Einstein tensor $G_{\alpha\beta} \equiv R_{\alpha\beta} - R g_{\alpha\beta}/2$. The last is constructed so as to have zero covariant divergence: $\nabla_\alpha G^{\alpha\beta} = 0$. These tensors enable us to look at the relationship between geometry and matter. Specifically, the Einstein tensor $G_{\alpha\beta}$, can be coupled via a constant κ to the energy-momentum tensor $T_{\alpha\beta}$ that describes properties of matter: $G_{\alpha\beta} = \kappa T_{\alpha\beta}$. These are Einstein's field equations. In the weak-field limit where $g_{00} \cong (1 + 2\phi/c^2)$ for a fluid of density ρ, Einstein's equations give back Poisson's equation $\nabla^2 \phi = 4\pi G \rho$. This presumes that the coupling constant is $\kappa = 8\pi G/c^4$, and shows that Einstein gravity contains Newton gravity. However, Einstein's field equations have only been rigorously tested in the solar system and the binary pulsar, where the gravitational field exists essentially in empty space or vacuum. In this case, $T_{\alpha\beta} = 0$ and the field equations $G_{\alpha\beta} = 0$ are equivalent to the simpler set

$$R_{\alpha\beta} = 0 \qquad (\alpha, \beta = 0 - 3) \quad . \tag{1.1}$$

These 10 relations serve in principle to determine the 10 $g_{\alpha\beta}$, and are the ones verified by observations.

Notwithstanding this, let us consider the full equations for a perfect isotropic fluid with density ρ and pressure p (i.e. there is no viscosity, and the pressure is equal in the 3 spatial directions). Then the energy-momentum tensor is $T_{\alpha\beta} = (p + \rho c^2) \mu_\alpha \mu_\beta - p g_{\alpha\beta}$ where μ_α, are the 4-velocities (see below). This is constructed so as to have zero divergence, and the equation of continuity and the equations of motion for the 3 spatial directions are derived from the 4 components of $\nabla_\alpha T^{\alpha\beta} = 0$. The covariant derivative here actually treats the metric tensor as a constant, so it is possible to add a term proportional to this to either the left-hand side or right-hand side of Einstein's equations. The former usage is traditional, so the full field equations are commonly written

$$R_{\alpha\beta} - \frac{Rg_{\alpha\beta}}{2} + \Lambda g_{\alpha\beta} = \frac{8\pi G}{c^4}\left[(p + \rho c^2)\mu_\alpha \mu_\beta - p g_{\alpha\beta}\right] \quad . \tag{1.2}$$

Here Λ is the cosmological constant, and its modulus is known to be small. It corresponds in the weak-field limit to a force per unit mass $|\Lambda|c^2 r/3$ which increases with radius r from the centre of (say) the solar system, but is not observed to significantly affect the orbits of the planets. However, it could be insignificant locally but significant globally, as implied by its dimensions (L^{-2}) In this regard, it is instructive to move the Λ term over to the other side of the field equations and incorporate it into $T_{\alpha\beta}$ as a "vacuum" contribution to the density and pressure:

$$\rho_\nu = \frac{\Lambda c^2}{8\pi G} \qquad p_\nu = \frac{\Lambda c^2}{8\pi G} \quad . \tag{1.3}$$

This "vacuum fluid" has the equation of state $p_\nu = -\rho_\nu c^2$, and while ρ_ν, is small by laboratory standards it could in principle be of the same order of magnitude as the material density of the galaxies ($10^{-29} - 10^{-31}$ gm cm^{-3}). Also, while $|\Lambda|$ is constrained by general relativity and observations of the present universe, there are arguments concerning the stability of the vacuum from quantum field theory which imply that it could have been larger in the early universe. But Λ (and G, c) are true constants in the original version of general relativity, so models of quantum vacuum transitions involve step-like phase changes (see e.g. Henriksen, Emslie and Wesson 1983). It should also be noted that while matter in the present universe has a pressure that is positive or close to zero ("dust"), there is in principle no reason why in the early universe or other exotic situations it cannot be taken negative. Indeed, any microscopic process which causes the particles of a fluid to attract each other can in a macroscopic way be described by $p < 0$ (the vacuum treated classically is a simple example). In fact, it is clear that p and ρ in general relativity are phenomenological, in the sense that they are labels for unexplained particle processes. It is also clear that the prime function of G and c is to dimensionally transpose matter labels such as p and ρ so that they match the geometrical objects of the theory.

The pressure and density are intimately connected to the motion of the fluid which they describe. This can be appreciated by looking at the general equation of motion, in the form derived by Raychaudhuri, and the continuity or conservation equation:

$$\begin{aligned} \ddot{R}\left(\frac{3}{R}\right) &= 2(w^2 - \sigma^2) - \frac{4\pi G}{c^2}(3p + \rho c^2) \\ \dot{\rho} c^2 &= -(p + \rho c^2)\frac{3\dot{R}}{R} \quad . \end{aligned} \tag{1.4}$$

Here R is the scale factor of a region of fluid with vorticity w, shear σ, and uniform pressure and density (see Ellis 1984: a dot denotes the total derivative with respect to time, and R should not be confused with the Ricci scalar introduced above and should not be taken as implying the existence of a physical boundary). From the first of (1.4) we see that the acceleration caused by a portion of the fluid depends on the combination $(3p+\rho c^2)$, so for mass to be attractive and positive we need $(3p+\rho c^2) > 0$. From the second of (1.4), we see that the rate of change of density depends on the combination $(p+\rho c^2)$, so for matter to be stable in some sense we need $(p+\rho c^2) > 0$. These inequalities, sometimes called the energy conditions, should not however be considered sacrosanct. Indeed, gravitational energy is a slippery concept in general relativity, and there are several alternative definitions of "mass" (Hayward 1994). These go beyond the traditional concepts of active gravitational mass as the agent which causes a gravitational field, passive gravitational mass as the agent which feels it, and inertial mass as the agent which measures energy content (Bonnor 1989). What the above shows is that in a fluid-dynamical context, $(3p + \rho c^2)$ is the gravitational energy density and $(p + \rho c^2)$ is the inertial energy density.

For a fluid which is homogeneous and isotropic ($\equiv$ uniform), without vorticity or shear, Einstein's equations reduce to 2 relations commonly called after Friedmann:

$$\begin{aligned} 8\pi G\rho &= \frac{3}{R^2}(kc^2 + \dot{R}^2) - \Lambda c^2 \quad , \\ \frac{8\pi G p}{c^2} &= -\frac{1}{R^2}(kc^2 + \dot{R}^2 + 2\ddot{R}R) + \Lambda c^2 \quad . \end{aligned} \tag{1.5}$$

Here $k = \pm 1,\ 0$ is the curvature constant which describes the departure of the 3D part of spacetime from flat Minkowski (specified by $g_{\alpha\beta} = \eta_{\alpha\beta} =$ diagonal +1, -1, -1, -1). There are many solutions of (1.5) which are more or less in agreement with cosmological observations. The simplest is the Einstein-de Sitter model. It has $k = 0$, $\Lambda = 0$, $p = 0$, $\rho = 1/6\pi G t^2$ and a scale factor $R(t)$ which grows as $t^{2/3}$. However, it requires about 2 orders of magnitude more matter to be present than in the visible galaxies, a topic we will return to in Sections 1.6 and 4.2. In general, solutions of (1.5) are called Friedmann-Robertson-Walker (FRW), where the last two names refer to the workers who derived the metric for these uniform cosmological models. This metric is commonly given in two different coordinate systems, whose justification has to do with whether one takes the global view wherein all directions in 3D space are treated the same, or the local view wherein quantities are measured from us as 'centre'. Noting that the radial coordinates r are different, the (3D) isotropic and non-isotropic forms of the metric are given by:

$$\begin{aligned} ds^2 &= c^2dt^2 - \frac{R^2(t)}{(1 + kr^2/4)^2}[dr^2 + r^2 d\Omega^2] \\ ds^2 &= c^2dt^2 - R^2(t)\left[\frac{dr^2}{(1 - kr^2)} + r^2 d\Omega^2\right] \quad . \end{aligned} \tag{1.6}$$

Here $d\Omega^2 \equiv d\theta^2 + \sin^2\theta d\phi^2$ defines the angular part of the metric in spherical polar coordinates. A photon which moves radially in the field described by (1.6) is defined by $ds = 0$ with $d\theta = d\phi = 0$. Using the second of (1.6) its (coordinate-defined) velocity is then

$$\frac{dr}{dt} = \pm \frac{c(1 - kr^2)^{1/2}}{R(t)} \quad . \tag{1.7}$$

Here the sign choice corresponds to whether the photon is moving towards or away from us. The important thing, though, is that the "speed" of the photon is *not* c.

This parameter, as noted in Section 1.2, is commonly regarded as defining an upper limit to the speed of propagation of causal effects. However, this interpretation is only true in the local, special-relativity limit. In the global, general-relativity case the size of causally-connected regions is defined by the concept of the horizon. An excellent account of this is given by its originator, Rindler (1977, p. 215). In the cosmological application, there are actually 2 kinds of horizon. An event horizon separates those galaxies we can see from those we cannot ever see even as $t \to \infty$; a particle horizon separates those galaxies we can see from those we cannot see now at $t = t_0 (\cong 2 \times 10^{10}\ yr)$. FRW models exist which have both kinds of horizon, one but not the other, or neither. A model in the latter category is that of Milne. (It has $k = -1$, $\Lambda = 0$, $p = 0$ and $R(t)$ proportional to t, and would solve the so-called horizon problem posed by the $3K$ microwave background did it not also have $\rho = 0$.) The distance to the particle horizon defines the size of that part of the universe which is in causal communication with us. The distance can be worked out quite simply for any k if we assume $\Lambda = p = 0$ (Weinberg 1972, p. 489). In terms of Hubble's parameter now ($H_0 \equiv \dot{R}_0/R_0$) and the deceleration parameter now ($q_0 = -\ddot{R}_0 R_0/\dot{R}_0^{\,2}$), the distances are given by:

$$
\begin{aligned}
d_{k=+1} &= \frac{c}{H_0(2q_0-1)^{1/2}}\cos^{-1}\left(\frac{1}{q_0}-1\right) \quad , \quad q_0 > \frac{1}{2} \\
d_{k=0} &= \frac{2c}{H_0} = 3ct_0 \quad , \quad q_0 = \frac{1}{2} \\
d_{k=-1} &= \frac{c}{H_0(1-2q_0)^{1/2}}\cosh^{-1}\left(\frac{1}{q_0}-1\right) \quad , \quad q_0 < \frac{1}{2} \quad .
\end{aligned}
\tag{1.8}
$$

Even for the middle case, the Einstein-de Sitter model with flat 3-space sections, the distance to the horizon is not ct_0. This confirms what was noted above, and shows that in relativity the purpose of c is merely to transpose a time to a length.

Particles with finite as opposed to zero rest masses move not along paths with $ds = 0$ but along paths with s a minimum. In particle physics with a special-relativity metric, the action principle for the motion of a particle with mass m is commonly written $\delta[\int m ds] = 0$. Assuming m = constant and replacing ds by its general relativity analog using $ds^2 = g_{\alpha\beta}dx^\alpha dx^\beta$, the variation leads to 4 equations of motion:

$$
\frac{du^\gamma}{ds} + \Gamma^\gamma_{\alpha\beta}u^\alpha u^\beta = 0 \quad . \tag{1.9}
$$

This is the geodesic equation, and its 4 components serve in principle to determine the 4-velocities $u^\gamma \equiv dx^\gamma/ds$ as functions of the coordinates. We note that, in addition to the *assumption* that m is constant, m does not appear in (1.9): general relativity is not a theory of forces but a theory of accelerations. In practice, (1.9) can only be solved algebraically for certain solutions of the field equations. The latter in vacuum are (1.1), and we note here that these can be obtained from an action via $\delta[\int R(-g)^{1/2}d^4x] = 0$. Here g is the determinant of the metric tensor, which with the conventional split of spacetime into time and space has signature (+ - - -) so g is negative. The field equations with matter can also be obtained from an action, but split into a geometrical part and a matter part. However, the split of a metric into time and space parts, and the split of the field equations into geometric and matter parts, are to a certain extent subjective.

1.4 Particle physics

This has evolved along different lines than gravitation, and while general relativity is monolithic, the standard model of particle physics is composite. Of relevance are the books by Ramond (1981), Griffiths (1987), and Collins, Martin and Squires (1989). The last is a good review of the connections between particle physics and cosmology, and also treats higher-dimensional theories of the types we will examine in subsequent sections. However, the present section is mainly concerned with standard 4D particle physics as based on Lagrangians, and the conceptual differences between gravitation and quantum theory.

The material is ordered by complexity: we consider the equations of Maxwell, Schrodinger, Klein-Gordon, Dirac, Proca and Yang-Mills; and then proceed to quantum chromodynamics and the standard model (including Glashow-Salam-Weinberg theory). As before, there is an emphasis on fundamental constants and the number of parameters required to make theory compatible with observation.

Classical electromagnetism is described by a 4-potential A_α and a 4-current J^α (covariant and contravariant quantities differ now by at most a sign). Then Maxwell's equations are contained in the tensor relations

$$\frac{\partial F^{\alpha\beta}}{\partial x^\alpha} = \frac{4\pi}{c} J^\beta \quad , \quad F_{\alpha\beta} \equiv \frac{\partial A^\beta}{\partial x^\alpha} - \frac{\partial A_\alpha}{\partial x^\beta} \quad , \tag{1.10}$$

and the identities

$$\frac{\partial F_{\alpha\beta}}{\partial x^\gamma} + \frac{\partial F_{\beta\gamma}}{\partial x^\alpha} + \frac{\partial F_{\gamma\alpha}}{\partial x^\beta} = 0 \tag{1.11}$$

implicit in the definition of the Faraday tensor $F_{\alpha\beta}$. However, Maxwell's equations may also be obtained by substituting the Lagrangian

$$L = -\frac{1}{16\pi} F^{\alpha\beta} F_{\alpha\beta} - \frac{1}{c} J^\alpha A_\alpha \tag{1.11}$$

in the Euler-Lagrange equations, which give (1.10). Strictly, L here is a Lagrangian density and has dimensions energy/volume, presuming we use the c.g.s./e.s.u. system of units. These units also imply that ϵ_0 does not appear (see Section 1.2). Thus c is the only constant that figures, in analogy with the original version of general relativity in which only G/c^4 figured (no cosmological constant). This is connected with the fact that these theories describe photons and gravitons with exactly zero rest mass.

Planck's constant $\hbar$ comes into the field theory of particles when the 3-momentum p and total energy E of a particle are replaced by space and time operators that act on a wave-function Ψ. Thus the prescriptions $p \to (\hbar/i)\nabla$ and $E \to (i\hbar)\partial/\partial t$ applied to the non-relativistic energy equation $p^2/2m + V = E$ (where m is rest mass and V is the potential energy) result in the Schrodinger equation

$$-\frac{\hbar^2}{2m}\nabla^2\Psi + V\Psi = i\hbar\frac{\partial\Psi}{\partial t} \quad . \tag{1.12}$$

The path Lagrangian for this is $L = T - V$ in general, which for a particle with charge q moving with a 3-velocity $dx/dt \ll c$ in an electromagnetic field is $L = (m/2)(dx/dt)^2 - (q/c)A_\alpha dx^\alpha/dt$. The path action for this is $S = \int_1^2 L dt$, where the integral is between two points. The variation $\delta S = 0$ gives the equations of motion of the particle between these two points, which in classical theory is a unique path. In quantum theory, there are non-unique paths, but the sum over paths $\Sigma exp(iS/\hbar)$ has the interpretation that the modulus squared is the probability that the particle goes from position 1 to 2. Clearly the phase $S/\hbar$ has to be dimensionless, and this is why $\hbar$ appears in the sum over paths. Instead of including it in the latter thing, however, we could instead use $\Sigma exp(iS)$ and redefine the Lagrangian to be

$$L = \frac{m}{2\hbar}\left(\frac{dx}{dt}\right)^2 - \frac{q}{c\hbar}A_\alpha\frac{dx^\alpha}{dt} \quad . \tag{1.13}$$

This has been pointed out by Hoyle and Narlikar (1974, p. 102; see also Ramond, 1981, p. 35). They go on to argue that since the second term in (1.13) contains another q implicit in A_α, it is the combination $q^2/\hbar$ that is important, and in it $\hbar$ can be absorbed into q^2. Also, in the first term in (1.13) it is the combination $m/\hbar$ that is important, and in it $\hbar$ can be absorbed into m. Thus the Lagrangian reduces back to the form given before.

A similar prescription to that above applied to the relativistic energy equation $E^2 - p^2c^2 = m^2c^4$ or $p^\alpha p_\alpha = m^2c^2$ for a freely-moving particle ($V = 0$) results in the Klein-Gordon equation

$$-\frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} + \nabla^2\phi = \left(\frac{mc}{\hbar}\right)^2\phi \quad . \tag{1.14}$$

Here ϕ is a single scalar field and the Lagrangian is

$$L = \frac{1}{2}\left[\left(\frac{1}{c}\frac{\partial\phi}{\partial t}\right)^2 - (\nabla\phi)^2\right] - \frac{1}{2}\left(\frac{mc}{\hbar}\right)^2\phi \quad . \tag{1.15}$$

Equations (1.14) and (1.15) describe a spin-0 particle in flat spacetime. We will consider the generalization to curved spacetime below.

Spin-1/2 particles were described in another equation formulated by Dirac, who 'factorized' the energy relation $p^\alpha p_\alpha = m^2c^2$ with the help of four 4×4 matrices γ^α. These latter are related to the metric tensor of Minkowski spacetime by the relation $\gamma^\alpha\gamma^\beta + \gamma^\beta\gamma^\alpha = 2\eta^{\alpha\beta}$. The Dirac equation is

$$i\hbar\gamma^\alpha\frac{\partial\Psi}{\partial x^\alpha} - mc\Psi = 0 \quad . \tag{1.16}$$

Here Ψ is a bi-spinor field, which can be thought of as a 4-element column matrix (though it is not a 4-vector) in which the upper two elements represent the two possible spin states of an electron while the lower two elements represent the two possible spin states of a positron. The Lagrangian is

$$L = i\hbar c\overline{\Psi}\gamma^\alpha\frac{\partial\Psi}{\partial x^\alpha} - mc^2\overline{\Psi}\Psi \quad . \tag{1.17}$$

Here $\overline{\Psi}$ is the adjoint spinor defined by $\overline{\Psi} \equiv \Psi^+\gamma^0$, where Ψ^+ is the usual Hermitian or transpose conjugate obtained by transposing Ψ from a column to a row matrix and complex-conjugating its elements. The Lagrangian (1.17) is for a free particle. It is invariant under the global gauge or phase transformation $\Psi \to e^{i\theta}\Psi$ (where θ is any real number), because $\overline{\Psi} \to e^{-i\theta}\overline{\Psi}$ and the exponentials cancel out in the combination $\overline{\Psi}\Psi$. But it is not invariant under the local gauge transformation $\Psi \to e^{i\theta(x)}\Psi$ which depends on location in spacetime. If the principle of local gauge invariance is desired, it is necessary to replace (1.17) by

$$L = i\hbar c\overline{\Psi}\gamma^\alpha\frac{\partial\Psi}{\partial x^\alpha} - mc^2\overline{\Psi}\Psi - q\overline{\Psi}\gamma^\alpha\Psi A_\alpha \quad . \tag{1.18}$$

Here A_α is a potential which we identify with electromagnetism and which changes under local gauge transformations according to $A_\alpha \to A_\alpha + \partial\lambda/\partial x^\alpha$ where $\lambda(x^\alpha)$ is a scalar function. In fact, we can say that the requirement of local gauge invariance for the Dirac Lagrangian (1.18) obliges the introduction of the field A_α typical of electromagnetism.

Actually the Lagrangian (1.18) should be even further extended by including a 'free' term for the gauge field. In this regard, the transformation $A_\alpha \to A_\alpha + \partial\lambda/\partial x^\alpha$ leaves $F_{\alpha\beta}$ unchanged, but not a term like $A^\alpha A_\alpha$. The appropriate term to add to (1.18) is therefore $(-1/16\pi)F^{\alpha\beta}F_{\alpha\beta}$, so the full Dirac Lagrangian is

$$L = i\hbar c\overline{\Psi}\gamma^\alpha\frac{\partial\Psi}{\partial x^\alpha} - mc^2\overline{\Psi}\Psi - \frac{1}{16\pi}F^{\alpha\beta}F_{\alpha\beta} - q\overline{\Psi}\gamma^\alpha\Psi A_\alpha \quad . \tag{1.19}$$

If we define a current density $J^\alpha \equiv cq(\overline{\Psi}\gamma^\alpha\Psi)$, the last two terms give back Maxwell's Lagrangian (1.11). The Lagrange density (1.19) describes electrons or positrons interacting with an electromagnetic field consisting of *massless* photons. However, a term like the one we just discarded ($A^\alpha A_\alpha$) may be acceptable in a theory of *massive* gauge particles. Indeed, a field derived from a vector potential A_α associated with a particle of finite rest mass m is described by the Proca equation

$$\frac{\partial F^{\alpha\beta}}{\partial x^\alpha} + \left(\frac{mc}{\hbar}\right)^2 A^\beta = 0 \quad . \tag{1.20}$$

This describes a spin-1 particle such as a massive photon, and can be obtained from the Lagrangian

$$L = -\frac{1}{16\pi} F^{\alpha\beta} F_{\alpha\beta} + \frac{1}{8\pi}\left(\frac{mc}{\hbar}\right)^2 A^\alpha A_\alpha \quad . \tag{1.21}$$

Again we see the combination $m/\hbar$, so $\hbar$ may be absorbed here if so desired as it has been elsewhere.

If we consider *two* 4-component Dirac fields, it can be shown that a locally gauge-invariant Lagrangian can only be obtained if we introduce three vector fields (A^1_α, A^2_α, A^3_α). These can be thought of as a kind of 3-vector A_α. It is also necessary to change the definition of $F_{\alpha\beta}$, used above. The 3 components of the new quantity ($F^1_{\alpha\beta}$, $F^2_{\alpha\beta}$, $F^3_{\alpha\beta}$) can again be thought of as a kind of vector, where now $F_{\alpha\beta} \equiv [\partial A_\beta/\partial x^\alpha - \partial A_\alpha/\partial x^\beta - (2q/\hbar c)(A_\alpha \times A_\beta)]$. Further, the three Pauli matrices (τ_1, τ_2, τ_3) can be regarded as a vector τ. Then with dot products between vectors defined in the usual way, the Lagrangian is

$$L = i\hbar c\overline{\Psi}\gamma^\alpha \frac{\partial \Psi}{\partial x^\alpha} - mc^2\overline{\Psi}\Psi - \frac{1}{16\pi} F^{\alpha\beta} \cdot F_{\alpha\beta} - (q\overline{\Psi}\gamma^\alpha \tau \Psi) \cdot A_\alpha \quad . \tag{1.22}$$

Here Ψ can be thought of as a column matrix with elements Ψ_1 and Ψ_2, each of which is a 4-component Dirac spinor. The latter still describe spin-1/2 particles of mass m (where we have assumed both particles to have the same mass for simplicity), and they interact with three gauge fields A^1_α, A^2_α, A^3_α which by gauge invariance must be massless. The kind of gauge invariance obeyed by (1.22) is actually more complex than that involving global and local phase transformations with $e^{i\theta}$ considered above. There Ψ was a single spinor, whereas here Ψ is a 2-spinor column matrix. This leads us to consider a 2×2 matrix which we take to be unitary ($U^+ U = 1$). In fact the first two terms in (1.22) are invariant under the global transformation $\Psi \to U\Psi$, because $\overline{\Psi} \to \overline{\Psi} U^+$ so the combination $\overline{\Psi}\Psi$ is invariant. Just as any complex number of modulus 1 can be written as $e^{i\theta}$ with θ real, any unitary matrix can be written $U = e^{1H}$ with H Hermitian ($H^+ = H$). Since H is a 2×2 matrix it involves 4 real numbers, say θ and a_1, a_2, a_3 which can be regarded as the components of a 3-vector a. As before, let τ be the 3-vector whose components are three 2×2 Pauli matrices, and let 1 stand for the 2×2 unit matrix. Then without loss of generality we can write $H = \theta 1 + \tau \cdot a$, so $U = e^{i\theta} e^{i\tau \cdot a}$. The first factor here is the old phase transformation. The second is a 2×2 unitary matrix which is special in that the determinant is actually 1. Thus $\Psi \to e^{i\tau\cdot a}\Psi$ is a global special-unitary 2-parameter, or $SU(2)$, transformation. It should be recalled that this global invariance only involves the first two terms of the Lagrangian (1.22), which resemble the Lagrangian (1.17) of Dirac. The passage to local invariance along lines similar to those considered above leads to the other terms in the Lagrangian (1.22) and was made by Yang and Mills.

The full Yang-Mills Lagrangian (1.22) is invariant under local $SU(2)$ gauge transformations, and leads to field equations that were originally supposed to describe two equal-mass spin-1/2 particles interacting with three massless spin-1 (vector) particles. In this form the theory is somewhat unrealistic, but still useful. For example, if we drop the first two terms in (1.22) we obtain a Lagrangian for the three gauge fields alone which leads to an interesting classical-type field theory that resembles Maxwell electrodynamics. This correspondence becomes clear if like before we define currents $J^\alpha \equiv cq(\overline{\Psi}\gamma^\alpha\tau\Psi)$, whereby the last two terms in (1.22) give a gauge-field Lagrangian

$$L = -\frac{1}{16\pi} F^{\alpha\beta} \cdot F_{\alpha\beta} - \frac{1}{c} J^{\alpha} \cdot A_{\alpha} \quad . \tag{1.23}$$

This closely resembles the Maxwell Lagrangian (1.11). But of course (1.23) gives rise to a considerably more complicated theory, solutions of which have been reviewed by Actor (1979). Some of these represent magnetic monopoles, which have not been observed. Some represent instantons and merons, which are hypothetical particles that tunnel between topologically distinct vacuum regions. Tunneling can in principle be important cosmologically. For example, Vilenkin (1982) has suggested that a certain type of instanton tunneling to de Sitter space from nothing can give birth to an inflationary universe. However, it is doubtful if the kinds of particles predicted by pure $SU(2)$ Yang-Mills theory will ever have practical applications. The real importance of this theory is that it showed it was feasible to use a symmetry group involving non-commuting 2×2 matrices to construct a non-Abelian gauge theory. This idea led to more successful theories, notably one for the strong interaction based on $SU(3)$ colour symmetry.

Quantum chromodynamics (QCD) is described by 3 coloured Dirac spinors that can be denoted Ψ_{red}, Ψ_{blue}, Ψ_{green} and 8 gauge fields given by a kind of 8-vector A_α. Each of Ψ_r, Ψ_b, Ψ_g is a 4-component Dirac spinor, and it is convenient to regard them as the elements of a column matrix Ψ. This describes the colour states of a massive spin-1/2 quark. The 8 components of A_α, are associated with the 8 Gell-Mann matrices (λ_{1-8}), which are the $SU(3)$ equivalents of the Pauli matrices of $SU(2)$, and describe massless spin-1 gluons. The Lagrangian for QCD can be constructed by adding together 3 Dirac Lagrangians like (1.17) above (one for each colour), insisting on local $SU(3)$ gauge invariance (which brings in the 8 gauge fields), and adding in a free gauge-field term (using $F_{\alpha\beta}$ as defined above for the original Yang-Mills theory). The complete Lagrangian is

$$L = i\hbar c \overline{\Psi} \gamma^{\alpha} \frac{\partial \Psi}{\partial x^{\alpha}} - mc^2 \overline{\Psi} \Psi - \frac{1}{16\pi} F^{\alpha\beta} \cdot F_{\alpha\beta} - (q \overline{\Psi} \gamma^{\alpha} \lambda \Psi) \cdot A_{\alpha} \quad . \tag{1.24}$$

This resembles (1.22) above. However, the electric charge of a quark needs to be a fraction of e in order to account for the common hadrons as quark composites. And particle physics is best described by 6 quarks with different flavours (d, u, s, c, b, t) and different masses m. This means we really need 6 versions of (1.24) with different masses. A gluon does not carry electric charge, but it does carry colour charge. This is unlike its analogue the photon in electrodynamics, allowing bound gluon states (glueballs) and making chromodynamics generally quite complicated.

We do not need to go into the intricacies of QCD, especially since good reviews are available (Ramond 1981; Llewellyn Smith 1983; Griffiths 1987; Collins, Martin and Squires 1989). But a couple of points related to charges and masses are relevant to our discussion. In the case of electrons interacting via photons, the Dirac Lagrangian and the fact that $\alpha \equiv e^2/\hbar c \cong 1/137$ is small allows perturbation analysis to be used to produce very accurate models. Indeed, quantum electrodynamics (QED) gives predictions that are in excellent agreement with experiment. However, the coupling parameter whose asymptotic value is the traditional fine-structure constant is actually energy or distance dependent. As mentioned in Section 1.2, this is commonly ascribed to vacuum polarization. Thus, a positive charge (say) surrounded by virtual electrons and positrons tends to attract the former and repel the latter. (Virtual particles do not obey Heisenberg's uncertainty relation and in modern quantum field theory the vacuum is regarded as full of them.) There is therefore a screening process, which means that the effective value of the embedded charge (and α) increases as the distance decreases. In analogy with QED, there is a similar process in QCD, but due to the different nature of the interaction the coupling parameter decreases as the distance decreases. This is the origin of asymptotic freedom, whose converse is that quarks in (say) a proton feel a strong restoring force if they move outwards and are in fact confined. In addition to the variable nature of coupling 'constants' and charges, the masses in QCD are also not what they appear to be. The m which appears in a Lagrangian like (1.24) is not really a given parameter, but is believed to arise from the spontaneous symmetry breaking which exists when a symmetry of the Lagrangian is not shared by the vacuum. Thus a manifestly symmetric Lagrangian with massless gauge-field particles can be rewritten in a less symmetric form by redefining the fields in terms of fluctuations about a particular ground state of the vacuum. This results in the gauge-field particles becoming massive and in the appearance of a massive scalar field or Higgs particle. In QCD. the quarks are initially taken to be massless, but if they have Yukawa-type couplings to the Higgs particle then they acquire masses. The Higgs mechanism in QCD, however, is really imported from the theory of the weak interaction, and has been mentioned here to underscore that the masses of the quarks are not really fundamental parameters.

The theory of the weak interaction was originally developed by Fermi as a way of accounting for beta decay, but is today mainly associated with Glashow, Weinberg and Salam who showed that it was possible to unify the weak and electromagnetic interactions (for reviews see Salam 1980 and Weinberg 1980). As it is formulated today, the theory of the weak interaction involves mediation by 3 very massive intermediate vector (spin-1) bosons, two of which ($W^{\pm}$) are electrically charged and one of which (Z^0) is neutral. These can be combined with the photon of electromagnetism via the symmetry group $SU(2) \otimes U(1)$, which is however spontaneously broken by the mechanism outlined in the preceding paragraph. Actually, the massive Z^0 and the massless photon are combinations of states that depend on a weak mixing angle θ_w, whose value is difficult to calculate from theory but is $\theta_w \cong 29^o$ from experiment. The theory of the weak interaction, like QED and QCD, involves a coupling parameter which is not constant.

What we have been discussing in the latter part of this section are parts of the standard model of particle physics, which symbolically unifies the electromagnetic, weak and strong interactions via the symmetry group $U(1) \otimes SU(2) \otimes SU(3)$. An appealing feature of this theory is that with increasing energy the electromagnetic coupling increases while the weak and strong couplings decrease, suggesting that they come together at some unifying energy. This, however, is not known: it is probably of order 10^{16} GeV, but could be as large as the Planck mass of order 10^{19} GeV (see Weinberg 1983; Llewellyn Smith 1983; Ellis 1983; Kibble 1983; Griffiths 1987, p. 77; Collins, Martin and Squires 1989, p. 159). Also, there are uncertainties in the theory, notably to do with the QCD sector where the numbers of colors and flavors are conventionally taken as 3 and 6 respectively but could be different. This means that while in the conventional model there are 6 quark masses and 6 lepton masses, there could be more. In fact, if we include couplings and other things, there are. at least 20 parameters in the theory (Ellis 1983). One might hope to reduce this by using a simple unifying group for $U(1)$, $SU(2)$ and $SU(3)$. but the minimal example of $SU(5)$ does not actually help much in this regard. And then there is the perennial question: What about gravity?

1.5 Kaluza-Klein theory

The idea that the world may have more than 4 dimensions is due to Kaluza (1921), who with a brilliant insight realized that a 5D manifold could be used to unify Einstein's theory of general relativity (Section 1.3) with Maxwell's theory of electromagnetism (Section 1.4). After some delay, Einstein endorsed the idea, but a major impetus was provided by Klein (1926). He made the connection to quantum theory by assuming that the extra dimension was microscopically small, with a size in fact connected via Planck's constant h to the magnitude of the electron charge e (Section 1.2). Despite its elegance, though, this version of Kaluza-Klein theory was largely eclipsed by the explosive development first of wave mechanics and then of quantum field theory. However, the development of particle physics led eventually to a resurgence of interest in higher-dimensional field theories as a means of unifying the long-range and short-range interactions of physics. Thus did Kaluza-Klein 5D theory lay the foundation for modern developments such as 11D supergravity and 10D superstrings (Section 1.6). In fact, there is some ambiguity in the scope of the phrase "Kaluza-

Klein theory". We will mainly use it to refer to a 5D field theory, but even in that context there are several versions of it. The literature is consequently enormous, but we can mention the conference proceedings edited by De Sabbata and Schmutzer (1983), Lee (1984) and Appelquist, Chodos and Freund (1987). A recent comprehensive review of all versions of Kaluza-Klein theory is the article by Overduin and Wesson (1997a). The latter includes a short account of what is referred to by different workers as non-compactified, induced-matter or space-time-matter theory. Since this is the subject of the following chapters, the present section will be restricted to a summary of the main features of traditional Kaluza-Klein theory.

This theory is essentially general relativity in 5D, but constrained by two conditions. Physically, both have the motivation of explaining why we perceive the 4 dimensions of spacetime and (apparently) do not see the fifth dimension. Mathematically, they are somewhat different, however. (a) The so-called 'cylinder' condition was introduced by Kaluza, and consists in setting all partial derivatives with respect to the fifth coordinate to zero. It is an extremely strong constraint that has to be applied at the outset of calculation. Its main virtue is that it reduces the algebraic complexity of the theory to a manageable level. (b) The condition of compactification was introduced by Klein, and consists in the assumption that the fifth dimension is not only small in size but has a closed topology (i.e. a circle if we are only considering one extra dimension). It is a constraint that may be applied retroactively to a solution. Its main virtue is that it introduces periodicity and allows one to use Fourier and other decompositions of the theory.

There are now 15 dimensionless potentials, which are the independent elements in a symmetric 5×5 metric tensor g_{AB} ($A,\ B = 0-4$: compare section 1.3). The first 4 coordinates are those of spacetime, while the extra one $x^4 = l$ (say) is sometimes referred to as the "internal" coordinate in applications to particle physics. In perfect analogy with general relativity, one can form a 5D Ricci tensor R_{AB}, a 5D Ricci scalar R and a 5D Einstein tensor $G_{AB} \equiv R_{AB} - Rg_{AB}/2$. The field equations would logically be expected to be $G_{AB} = kT_{AB}$ with some appropriate coupling constant k and a 5D energy-momentum tensor. But the latter is unknown, so from the time of Kaluza and Klein onward much work has been done with the 'vacuum' or 'empty' form of the field equations $G_{AB} = 0$. Equivalently, the defining equations are

$$R_{AB} = 0 \qquad (A,\ B = 0-4) \quad . \tag{1.25}$$

These 15 relations serve to determine the 15 g_{AB}, at least in principle.

In practice, this is impossible without some starting assumption about g_{AB}. This is usually connected with the physical situation being investigated. In gravitational problems, an assumption about $g_{AB} = g_{AB}(x^c)$ is commonly called a choice of coordinates, while in particle physics it is commonly called a choice of gauge. We will meet numerous concrete examples later, where given the functional form of $g_{AB}(x^c)$ we will calculate the 5D analogs of the Christoffel symbols Γ^C_{AB} which then give the components of R_{AB} (Chapters 2-4). Kaluza was interested in electromagnetism, and realized that g_{ab} can be expressed in a form that involves the 4-potential A_α that figures in Maxwell's theory. He adopted the cylinder condition noted above, but also put $g_{44} =$ constant. We will do a general analysis of the electromagnetic problem later (Chapter 5), but here we look at an intermediate case where $g_{AB} = g_{AB}(x^\alpha)$, $g_{44} = -\Phi^2(x^\alpha)$. This illustrates well the scope of Kaluza-Klein theory, and has been worked on by many people, including Jordan (1947, 1955). Bergmann (1948), Thiry (1948), Lessner (1982), and Liu and Wesson (1997). The coordinates or gauge are chosen so as to write the 5D metric tensor in the form

$$g_{AB} = \begin{bmatrix} (g_{\alpha\beta} - \kappa^2\Phi^2 A_\alpha A_\beta) & -\kappa\Phi^2 A_\alpha \\ -\kappa\Phi^2 A_\beta & -\Phi^2 \end{bmatrix} \quad , \tag{1.26}$$

where κ is a coupling constant. Then the field equations (1.25) reduce to

$$
\begin{aligned}
G_{\alpha\beta} &= \frac{\kappa^2\Phi^2}{2}T_{\alpha\beta} - \frac{1}{\Phi}\Big(\nabla_\alpha\nabla_\beta\Phi - g_{\alpha\beta}\Box\Phi\Big) \\
\nabla^\alpha F_{\alpha\beta} &= -3\frac{\nabla^\alpha\Phi}{\Phi}F_{\alpha\beta} \\
\Box\Phi &= -\frac{\kappa^2\Phi^3}{4}F_{\alpha\beta}F^{\alpha\beta} \quad . \qquad (1.27)
\end{aligned}
$$

Here $G_{\alpha\beta}$ and $F_{\alpha\beta}$ are the usual 4D Einstein and Faraday tensors (see sections 1.3 and 1.4 respectively), and $T_{\alpha\beta}$, is the energy-momentum tensor for an electromagnetic field given by $T_{\alpha\beta} = (g_{\alpha\beta}F_{\gamma\delta}F^{\gamma\delta}/4 - F^{\gamma}_{\alpha}F_{\beta\gamma})/2$. Also $\Box \equiv g^{\alpha\beta}\nabla_\alpha\nabla_\beta$ is the wave operator, and the summation convention is in effect. Therefore we recognize the middle member of (1.27) as the 4 equations of electromagnetism modified by a function, which by the last member of (1.27) can be thought of as depending on a wave-like scalar field. The first member of (1.27) gives back the 10 Einstein equations of 4D general relativity, but with a right-hand side which in some sense represents energy and momentum that are effectively derived from the fifth dimension. In short, Kaluza-Klein theory is in general a unified account of gravity, electromagnetism and a scalar field.

Kaluza's case $g_{44} = -\Phi^2 = -1$ together with the identification $\kappa = \Big(16\pi G/c^4\Big)^{1/2}$ makes (1.27) read

$$
\begin{aligned}
G_{\alpha\beta} &= \frac{8\pi G}{c^4}T_{\alpha\beta} \\
\nabla^\alpha F_{\alpha\beta} &= 0 \quad . \qquad (1.28)
\end{aligned}
$$

These are of course the straight Einstein and Maxwell equations in 4D, but derived from vacuum in 5D, a consequence which is sometimes referred to as the Kaluza-Klein "miracle". However, these relations involve by (1.27) the choice of electromagnetic gauge $F_{\alpha\beta}F^{\alpha\beta} = O$ and have no contribution from the scalar field. The latter could well be important, particularly in application to particle physics. In the language of that subject, the field equations (1.25) of Kaluza-Klein theory describe a spin-2 graviton, a spin-1 photon and a spin-0 boson which is thought to be connected with how particles acquire mass. The field equations can also be derived from a 5D action $\delta\Big[\int R(-g)^{1/2}d^5x\Big] = 0$, in a way analogous to what happens in 4D Einstein theory.

It is also possible to put Kaluza-Klein theory into formal correspondence with other 4D theories, notably the Brans-Dicke scalar-tensor theory (see Overduin and Wesson 1997a). This theory is sometimes cast in a form where the scalar field is effectively disguised by putting the functional dependence into G, the gravitational 'constant'. In this regard it belongs to a class of 4D theories, which includes ones by Dirac, Hoyle and Narlikar and Canuto et al., where the constants are allowed to vary with cosmic time (see Wesson 1978 and Barbour and Pfister 1995 for reviews). However, it should be stated with strength that Kaluza-Klein theory is essentially 5D, and trying to cast it into 4D form is technically awkward. It should also be noted that the reasons for treating 4D fundamental constants in this way are conceptually obscure.

1.6 Supergravity and superstrings

These are based on the idea of supersymmetry, wherein each boson (integral spin) is matched with a fermion (half integral spin). Thus the particle which is presumed to mediate classical gravity (the graviton) has a partner (the gravitino). This kind of symmetry is natural, insofar as particle physics needs to account for both bosonic and fermionic matter fields. But it is also attractive because it leads to a cancellation of the enormous zero-point fields which otherwise exist but whose energy density is not manifested in the curvature of space (this is related to the so-called cosmological constant problem, which is discussed elsewhere). The literature on supergravity and superstrings is diverse, but we can mention the review articles by Witten (1981) and Duff (1996); and the books by West (1986) and Green, Schwan and Witten (1987). The status of the electromagnetic zero-point field has been discussed by Wesson (1991). There is an obvious connection between 5D Kaluza-Klein theory, 11D supergravity and 10D superstrings. But while the former is more-or-less worked out, the latter are still in a state of development with an uncertain prognosis where it comes to their relevance to the real world. For this reason, and also because supersymmetry lies outside the scope of the rest of this work, we will content ourselves with a short history.

Supersymmetric gravity or supergravity began life as a 4D theory in 1976 but quickly made the jump to higher dimensions ("Kaluza-Klein supergravity"). It was particularly successful in 11D, for three principal reasons. First, Nahm showed that 11 was the maximum number of dimensions consistent with a single graviton (and an upper limit of two on particle spin). This was followed by Witten's proof that 11 was also the minimum number of dimensions required for a Kaluza-Klein theory to unify all the forces in the standard model of particle physics (i.e. to contain the gauge groups of the strong $SU(3)$ and electroweak $SU(2) \otimes U(1)$ interactions). The combination of supersymmetry with Kaluza-Klein theory thus appeared to uniquely fix the dimensionality of the world. Second, whereas in lower dimensions one had to choose between several possible configurations for the matter fields, Cremmer et al. demonstrated in 1978 that in 11D there is a single choice consistent with the requirements of supersymmetry (in particular, that there be equal numbers of Bose and Fermi degrees of freedom). In other words, while a higher-dimensional energy-momentum tensor was still required, its form at least appeared somewhat natural. Third, Freund and Rubin showed in 1980 that compactification of the 11D model could occur in only two ways: to 7 or 4 compact dimensions, leaving 4 (or 7, respectively) macroscopic ones. Not only did 11D spacetime appear to be specially favored for unification, but it also split perfectly to produce the observed 4D world. (The other possibility, of a macroscopic 7D world, could however not be ruled out, and in fact at least one such model was constructed as well.) Buoyed by these three successes, 11D supergravity appeared set by the mid-1980s as a leading candidate for the hoped-for "theory of everything".

Unfortunately, certain difficulties have dampened this initial enthusiasm. For example, the compact manifolds originally envisioned by Witten (those containing the standard model) turn out not to generate quarks or leptons, and to be incompatible with supersymmetry. Their most successful replacements are the 7-sphere and the "squashed" 7-sphere, described respectively by the symmetry groups $SO(8)$ and $SO(5) \otimes SU(2)$. But these groups do not contain the minimum symmetry requirements of the standard model $[SU(3) \otimes SU(2) \otimes U(1)]$. This is commonly rectified by adding matter-related fields, the "composite gauge fields", to the 11D Lagrangian. Another problem is that it is very difficult to build chirality (necessary for a realistic fermion model) into an 11D theory. A variety of remedies have been proposed for this, including the common one of adding even more gauge fields, but none has been universally accepted. It should also be mentioned that supergravity theory is marred by a large cosmological constant in 4D, which is difficult to remove even by fine-tuning. Finally, quantization of the theory inevitably leads to anomalies.

Some of these difficulties can be eased by descending to 10 dimensions: chirality is easier to obtain, and many of the anomalies disappear. However, the introduction of chiral fermions leads to new kinds of anomalies. And the primary benefit of the 11D theory - its uniqueness - is lost: 10D is not specially favored, and the theory does not break down naturally into 4 macroscopic and 6 compact dimensions. (One can still find solutions in which this happens, but there is no reason why they should be preferred.) In fact, most 10D supergravity models not only require *ad hoc* higher-dimensional matter fields to ensure proper compactification, but entirely ignore gauge fields arising from the Kaluza-Klein mechanism (i.e. from symmetries of the compact manifold). A theory which requires all gauge fields to be effectively put in by hand can hardly be considered natural.

A breakthrough in solving the uniqueness and anomaly problems of 10D theory occured when Green and Schwarz and Gross et al. showed that there were 2 (and only 2) 10D supergravity models in which all anomalies could be made to vanish: those based on the groups $SO(32)$ and $E_8 \otimes E_8$, respectively. Once again, extra terms (known as Chapline-Manton terms) had to be added to the higher-dimensional Lagrangian. This time, however, the addition was not completely arbitrary; the extra terms were those which would appear anyway if the theory were a low-energy approximation to certain kinds of supersymmetric string theory.

Supersymmetric generalizations of strings, or superstrings, are far from being understood. However, they have some remarkable virtues. For example, they retain the appeal of strings, wherein a point particle is replaced by an extended structure, which opens up the possibility of an anomaly-free approach to quantum gravity. (They do this while avoiding the generic prediction of tachyons, which plagued the old string theories.) Also, it is possible to make connections between certain superstring states and extreme black holes. (This may help resolve the problem of what happens to the information swallowed by classical singularities, which has been long standing in general relativity.) It is true that, for a while, there was thought to be something of a uniqueness problem for 10D superstrings, in that the groups $SO(32)$ and $E_8 \otimes E_8$ admit five different string theories between them. But this difficulty was addressed by Witten, who showed that it is possible to view these five theories as aspects of a single underlying theory, now known as M-theory (for "Membrane"). The low-energy limit of this new theory, furthermore, turns out to be 11D supergravity. So it appears that the preferred dimensionality of spacetime may after all be 11, at least in regard to higher-dimensional theories which are compactified.

Supersymmetric particles such as gravitinos and neutralinos, if they exist, could provide the dark or hidden matter necessary to explain the dynamics of galaxies and bring cosmological observations into line with the simplest 4D cosmological models (see Section 1.3). However, such 'dark' matter is probably not completely dark, because the particles concerned are unstable to decay in realistic (non-minimal) supersymmetric theories, and will contribute photons to the intergalactic radiation field. Observations of the latter can be used to constrain supersymmetric weakly interacting massive particles (WIMPS). Thus gravitinos and neutralinos are viable dark-matter candidates if they have decay lifetimes greater than of order 10^{11} yr and 10^{9} yr respectively (Overduin and Wesson 1997b). In this regard, they are favored over non-supersymmetric candidates such as massive neutrinos, axions and a possible decaying vacuum (Overduin and Wesson 1997c, 1992). There are other candidates, but clearly the identification of dark matter is an important way of testing supersymmetry.

1.7 Conclusion

This chapter has presented a potted account of theoretical physics as it exists at the present. We have learned certain things, namely: that fundamental constants are not (Section 1.2); that general relativity describes gravity excellently in curved 4D space (Section 1.3); that particle physics works well as a composite theory in flat 4D space (Section 1.4); that Kaluza-Klein theory in its original version unifies gravity and electromagnetism in curved 5D space (Section 1.5); and that supergravity and superstrings provide possible routes to new physics in 11D and 10D. So, where do we go from here?

There is no consensus answer to this, but let us consider the following line of reasoning. Physics is a *description* of the world as *we* perceive it (Eddington). In order to give a logical and coherent account of the maximum number of physical phenomena, we should presumably use the most advanced mathematical techniques. For the last century through to now, this implies that we should use geometry (Einstein, Riemann). The field equations of general relativity have no mathematical constraint as regards the number of dimensions in which they should be applied the choice follows from physics and depends on what we wish to explain. Also, there are certain ways of embedding lower-dimensional spaces with complicated structure in higher-dimensional spaces with simple structure, including flat ones (Campbell, Eisenhart: see the next chapter). So the question of how we can best describe gravity and particle physics is to a certain extent a question of algebraic technology. Now we might expect that the many quantum properties of elementary particles should be described by a space with a large number of dimensions. However, the *classical* properties of matter should be able to be handled by a space with a moderate number of dimensions.

The rest of this treatise is a compilation of (mainly technical) results which demonstrates this view. It will be seen that properties of matter such as the density and pressure of a fluid, as well as the rest mass and electric charge of a particle, can be derived from 5D geometry. This may sound surprising, but there are important differences between what we do now and what others have done before. The theory we will be working with is obviously not Einstein general relativity, since it is not 4D but 5D in nature. But it is not Kaluza-Klein either, because we do not invoke the hobbling cylinder condition typical of that theory, preferring instead to examine an unrestricted and rich 5D algebra. What we do in the following chapters also differs from previous work in that we do not need an explicit energy-momentum tensor: it will be seen that matter can be derived from geometry.

Bibliography

[1] Actor, A. 1979, Rev. Mod. Phys. 51, 461.

[2] Appelquist, T., Chodos, A., Freund, P.G.O. 1987, Modem Kaluza-Klein Theories (Addison-Wesley, Menlo Park).

[3] Barbour, J.B., Pfister, H. (eds.) 1995, Mach's Principle: From Newton's Bucket to Quantum Gravity (Birkhauser, Boston).

[4] Barrow, J.D. 1981, Quart. J. Roy. Astron. Soc. 22, 388.

[5] Barrow, J.D. 1983, Quart. J. Roy. Astron. Soc. 24, 24.

[6] Barrow, J.D., Tipler, F.J. 1986, The Anthropic Principle (Oxford U. Press, New York).

[7] Bekenstein, J.D. 1979, Comm. Astrophys. 8, 89.

[8] Bergmann, P. 1948, Ann. Math. (Princeton) 49, 255.

[9] Bonnor, W.B. 1989, Gen. Rel. Grav. 2l, 1143.

[10] Bridgman, P.W. 1922, Dimensional Analysis (Yale U. Press, New Haven).

[11] Bunge, M. 1971 in Problems in the Foundations of Physics (ed. Bunge, M., Springer, Berlin). 1.

[12] Carr, B.J., Rees, M.J. 1979, Nature 278, 605.

[13] Collins, P.D.B., Martin, A.D., Squires, E.J. 1989, Particle Physics and Cosmology (Wiley, New York).

[14] De Sabbata, V., Schmutzer, E.(eds.) 1983, Unified Field Theories of More than 4 Dimensions Including Exact Solutions (World Scientific, Singapore).

[15] Desloge, E.A. 1984, Am. J. Phys. 52, 312.

[16] Dicke, R.H. 1962, Phys. Rev. 125, 2163.

[17] Duff, M.J. 1996, Int. J. Mod. Phys. A 11, 5623.

[18] Eddington, A.S. 1929, Science and the Unseen World (MacMillan, New York).

[19] Eddington, A.S. 1935, New Pathways in Science (Cambridge U. Press, Cambridge).

[20] Eddington, A.S. 1939, The Philosophy of Physical Science (Cambridge U. Press, Cambridge).

[21] Ellis, G.F.R. 1984, Ann. Rev. Astron. Astrophys. 22, 157.

[22] Ellis, J. 1983, Phil. Trans. Roy. Soc. (London) A 310, 279.

[23] Green, M.B., Schwarz, J.H., Witten, E. 1987, Superstring Theory (Cambridge U. Press, Cambridge).

[24] Griffiths, D. 1987, Introduction to Elementary Particles (Harper and Row, New York).

[25] Hayward, S.A. 1994, Phys. Rev. D 49, 831.

[26] Henriksen, R.N., Emslie, A.G., Wesson, P.S. 1983, Phys. Rev. D 27, 1219.

[27] Hoyle, F., Narlikar, J.V. 1974, Action at a Distance in Physics and Cosmology (Freeman, San Francisco).

[28] Jackson, J.D. 1975, Classical Electrodynamics (2nd ed., Wiley, New York).

[29] Jammer, M. 1961, Concepts of Mass in Classical and Modem Physics (Harvard U. Press, Cambridge).

[30] Jeffreys, H. 1973, Scientific Inference (3rd ed., Cambridge U. Press, Cambridge).

[31] Jordan, P. 1947, Ann. Phys. (Leipzig) 1, 219.

[32] Jordan, P. 1955, Schwerkraft und Weltall (Vieweg, Braunschweig).

[33] Kaluza, T. 1921, Sitz. Preuss. Akad. Wiss. 33, 966.

[34] Kibble, T.W.B. 1983, Phil. Trans. Roy. Soc. (London) A 310, 293.

[35] Klein, O. 1926, Z. Phys. 37, 895.

[36] Lee, H.C. (ed.) 1984, An Introduction to Kaluza-Klein Theories (World Scientific, Singapore).

[37] Lessner, G. 1982, Phys. Rev. D 25, 3202.

[38] Liu, H., Wesson, P.S. 1997, Class. Quant. Grav. l4, 1651.

[39] Llewellyn Smith, C.H. 1983, Phil. Trans. Roy Soc. (London) A 310, 253.

[40] McCrea, W.H. 1986, Quart. J. Roy. Astron. Soc. 27, 137.

[41] McCrea, W.H., Rees, M.J. (eds.) 1983, Phil Trans. Roy. Soc.(London) A 310, 209.

[42] Misner, C.W., Thorne, K.S., Wheeler, J.A. 1973, Gravitation (Freeman, San Francisco).

[43] Nodvik, J.S. 1985, Phys. Rev. Lett. 55, 2519.

[44] Overduin, J.M., Wesson, P.S. 1992, Vistas Astron. 35, 439.

[45] Overduin, J.M., Wesson, P.S. 1997a. Phys. Reports 283, 303.

[46] Overduin, J.M., Wesson, P.S. 1997b. Astrophys. J. 480, 470.

[47] Overduin, J.M., Wesson, P.S. 1997c, Astrophys. J. 483, 77.

[48] Petley, B.W. 1985, The Fundamental Constants and the Frontier of Measurement (Hilger, Bristol).

[49] Ramond, P. 1981, Field Theory (Benjamin-Cummings, Reading).

[50] Rindler, W. 1977, Essential Relativity (2nd ed., Springer, Berlin).

[51] Salam. A. 1980, Rev. Mod. Phys. 52, 525.

[52] Smith, K.F. 1983, Phil. Trans. Roy. Soc. (London) A 310, 215.

[53] Thiry, Y. 1948, Com. Rend. Acad. Sci. (Paris) 226, 216.

[54] Vilenkin, A. 1982, Phys. Lett. B 117, 25.

[55] Weinberg, S. 1972, Gravitation and Cosmology (Wiley, New York).

[56] Weinberg, S. 1980, Rev. Mod. Phys. 52, 515.

[57] Weinberg. S . 1983, Phil. Trans. Roy. Soc. (London) A 310, 249.

[58] Wesson, P.S. 1978, Cosmology and Geophysics (Oxford U. Press-Hilger, New York).

[59] Wesson, P.S. 1983, Astron. Astrophys. 119, 313.

[60] Wesson, P.S. 1991, Astrophys J. 378, 466.

[61] Wesson, P.S. 1992, Space Science Rev. 59, 365.

[62] West, P. 1986, Introduction to Supersymmetry and Supergravity (World Scientific, Singapore).

[63] Will, C.M. 1993, Theory and Experiment in Gravitational Physics (Cambridge U. Press, Cambridge).

[64] Witten, E. 1981, Nucl. Phys. B 186, 412.

Chapter 2

Induced-Matter Theory

"To make physics, the geometry should bite"

(John Wheeler, Princeton, 1984)

2.1 Introduction

We import from the preceding chapter two important and connected ideas. First, as realized by several workers, the so-called fundamental constants like c, G and h have as their main purpose the transposition of physical dimensions. Thus, a mass can be regarded as a length; and physical quantities such as density and pressure can be regarded as having the same dimensions as the geometrical quantities that figure in general relativity. Second, physical quantities should be given a geometric interpretation, as envisaged by many people through time, including Einstein who wished to transmute the "base wood" of physics to the "marble" of geometry. An early attempt at this was made by Kaluza and Klein, who extended general relativity from 4 to 5 dimensions, but also applied severe restrictions to the geometry (the conditions of cylindricity and compactification). In this chapter, we will draw together results which have appeared in recent years which show that it is possible to interpret most properties of matter as the result of 5D Riemannian geometry, where however the latter allows dependence on the fifth coordinate and does not make assumptions about the topology of the fifth dimension.

This induced-matter theory has seen most work in 3 areas: (a) The case of uniform cosmological models is easiest to treat because of the high degree of symmetry involved, and is very instructive. (b) The soliton case is more complicated, but important because 5D solitons are the analogs of isolated 4D masses, and the 5D class of soliton solutions contains the unique 4D Schwarzschild solution. (c) The case of neutral matter can be treated quite generally, and lays the foundation for many applications where electromagnetic effects are not involved. After an outline of geometric feasibility (Section 2.2) we will give the main theoretical results in each of the aforementioned areas (Sections 2.3, 2.4, 2.5). We defer the main observational implications to later chapters. Our conclusion (Section 2.6) will be that *one* extra dimension is enough to explain the phenomenological properties of *classical* matter.

From here on, we will absorb the fundamental constants c, G and h via a choice of units that renders their magnitudes unity. We will use the metric signature with diagonal $= (+ - - - \pm)$, where the last choice will be seen to depend on the physical application and not cause any problem with causality. Also, we will label 5D quantities with upper-case Latin letters ($A = 0 - 4$) and 4D quantities with lower-case Greek letters ($\alpha = 0 - 3$). If there is a chance of confusion between the 4D part of a 5D quantity and the 4D quantity as conventionally defined, we will use a hat to denote the former and the straight symbol to denote the latter.

2.2 A 5D embedding for 4D matter

The 5D field equations for apparent vacuum in terms of the Ricci tensor are

$$R_{AB} = 0 \quad . \tag{2.1}$$

Equivalently, in terms of the 5D Ricci scalar and the 5D Einstein tensor $G_{AB} \equiv R_{AB} - Rg_{AB}/2$, they are

$$G_{AB} = 0 \quad . \tag{2.2}$$

By contrast, the 4D field equations with matter are given by Einstein's relations of general relativity:

$$G_{\alpha\beta} = 8\pi T_{\alpha\beta} \quad . \tag{2.3}$$

The central thesis of induced-matter theory is that (2.3) are a subset of (2.2) with an effective or induced 4D energy-momentum tensor $T_{\alpha\beta}$ which contains the classical properties of matter.

That this is so will become apparent below when we treat several cases suggested by physics. However, it is also possible to approach the subject through algebra; and while results in the latter field were subsequent, they are general and can be summarized here. Thus, it is a direct consequence of a little-known theorem by Campbell that any analytic N-dimensional Riemannian manifold can be locally embedded in an (N+1)-dimensional Ricci-flat ($R_{AB} = 0$) Riemannian manifold (Romero, Tavakol and Zalaletdinov 1996). This is of great importance for establishing the generality of the proposal that 4D field equations with sources can be locally embedded in 5D field equations without sources. And it can be used to study lower-dimensional ($N < 4$) gravity, which may be easier to quantize than general relativity (Rippl, Romero and Tavakol 1995). It can also be employed to find new classes of 5D solutions (Lidsey et al. 1997). Some of the latter have the remarkable property that they are 5D flat but contain 4D subspaces that are curved and correspond to known physical situations (Wesson 1994; Abolghasem, Coley and McManus 1996). The latter do not, though, include the 4D Schwarszchild solution, which can only be embedded in a flat manifold with $N \geq 6$ (Schouten and Struik 1921; Tangherlini 1963; for general results in embeddings see Campbell 1926; Eisenhart 1949; Kramer et al. 1980). However, the principle is clear: curved 4D physics can be embedded in curved or flat 5D geometry, and we proceed to study 3 prime cases of this.

2.3 The cosmological case

There are many exact solutions known of (2.1) that are of cosmological type, meaning that the metric resembles that of Robertson-Walker and the dynamics is governed by equations like those of Friedmann (see Section 1.3). However, most of these do not involve dependence on the extra coordinate l and are from the induced-matter viewpoint very restricted. Thus while we will use one of these solutions below, we will concentrate on the much more significant solutions of Ponce de Leon (1988). He found several classes of exact solutions of (2.1) whose metrics are separable and reduce to the standard 4D RW ones on the hypersurfaces l = constants. The induced matter and other properties associated with the most physical class of Ponce de Leon solutions were worked out by Wesson (1992a). Since then, many other cosmological solutions and their associated matter properties have been derived by various workers (see, e.g., Chatterjee and Sil 1993; Chatterjee, Panigrahi and Banerjee 1994; Liu and Wesson 1994; Liu and Mashhoon 1995; Billyard and Wesson 1996). In what follows, we will illustrate the transition from the 5D equations (2.1), (2.2) for apparent vacuum to the 4D equations (2.3) with matter, by using simple but realistic solutions.

It is convenient to consider a 5D metric with interval given by

$$dS^2 = e^{\nu} dt^2 - e^{\omega}(dr^2 + r^2 d\Omega^2) - e^{\mu} dl^2 \quad . \tag{2.4}$$

Here the time coordinate $x^0 = t$ and the space coordinates $x^{123} = r\theta\phi(d\Omega^2 \equiv d\theta^2 + sin^2\theta d\phi^2)$ have been augmented by the new coordinate $x^4 = l$. The metric coefficients ν, ω, and μ will depend in general on both t and l, partial derivatives with respect to which will be denoted by an overdot and an asterisk, respectively. Components of the Einstein tensor in mixed form are:

$$\begin{aligned}
G_0^0 &= -e^{-\nu}\left(-\frac{3\dot{\omega}^2}{4} - \frac{3\dot{\omega}\dot{\mu}}{4}\right) + e^{-\mu}\left(\frac{3\omega^{**}}{2} + \frac{3\omega^{*2}}{2} - \frac{3\mu^*\omega^*}{4}\right) \\
G_4^0 &= e^{-\nu}\left(\frac{3\omega^{**}}{2} + \frac{3\dot{\omega}\omega^*}{4} - \frac{3\dot{\omega}\nu^*}{4} - \frac{3\omega^*\dot{\mu}}{4}\right) \\
G_1^1 &= G_2^2 = G_3^3 = -e^{-\nu}\left(\ddot{\omega} + \frac{3\dot{\omega}^2}{4} + \frac{\ddot{\mu}}{2} + \frac{\dot{\mu}^2}{4} + \frac{\dot{\omega}\dot{\mu}}{2} - \frac{\dot{\nu}\dot{\omega}}{2} - \frac{\dot{\nu}\dot{\mu}}{4}\right) \\
&\quad + e^{-\mu}\left(\omega^{**} + \frac{3\omega^{*2}}{4} + \frac{\nu^{**}}{2} + \frac{\nu^{*2}}{4} + \frac{\omega^*\nu^*}{2} - \frac{\mu^*\omega^*}{2} - \frac{\nu^*\mu^*}{4}\right) \\
G_4^4 &= -e^{-\nu}\left(\frac{3\ddot{\omega}}{2} + \frac{3\dot{\omega}^2}{2} - \frac{3\dot{\nu}\dot{\omega}}{4}\right) + e^{-\mu}\left(\frac{3\omega^{*2}}{4} + \frac{3\omega^*\nu^*}{4}\right) \quad .
\end{aligned} \tag{2.5}$$

These are 5D components. We wish to match the terms in (2.5) with the components of the usual 4D perfect-fluid energy-momentum tensor. This is $T_{\alpha\beta} = (p+\rho)\mu_\alpha\mu_\beta - pg_{\alpha\beta}$, where $\mu^\alpha \equiv dx^\alpha/ds$, and for our case has components $T_0^0 = \rho$, $T_1^1 = -p$ for the density and pressure. Following the philosophy outlined above, we simply identify the new terms (due to the fifth dimension) in G_0^0 with ρ, and the new terms in G_1^1 with p. Then, collecting terms which depend on the new metric coefficient μ or derivatives with respect to the new coordinate l, we define

$$\begin{aligned}
8\pi\rho &\equiv -\frac{3}{4}e^{-\nu}\dot{\omega}\dot{\mu} + \frac{3}{2}e^{-\mu}\left(\omega^{**}\omega^{*2} - \frac{\mu^*\omega^*}{2}\right) \\
8\pi p &\equiv e^{-\nu}\left(\frac{\ddot{\mu}}{2} + \frac{\dot{\mu}^2}{4} + \frac{\dot{\omega}\dot{\mu}}{2} - \frac{\dot{\nu}\dot{\mu}}{4}\right) \\
&\quad - e^{-\mu}\left(\omega^{**} + \frac{3\omega^{*2}}{4} + \frac{\nu^{**}}{2} + \frac{\nu^{*2}}{4} + \frac{\omega^*\nu^*}{2} - \frac{\mu^*\omega^*}{2} - \frac{\nu^*\mu^*}{4}\right) \quad .
\end{aligned} \tag{2.6}$$

These are suggested identifications for 4D properties of matter in terms of 5D properties of geometry.

To see if they make physical sense to this point, we combine (2.6) and (2.5) with the field equations $G_B^A = 0$ of (2.2). There comes

$$\begin{aligned}
G_0^0 &= -\frac{3}{4}e^{-\nu}\dot{\omega}^2 + 8\pi\rho = 0 \\
G_4^0 &= e^{-\nu}\left(\frac{3\omega^{**}}{2} + \frac{3\dot{\omega}\omega^*}{4} - \frac{3\dot{\omega}\nu^*}{4} - \frac{3\omega^*\dot{\mu}}{4}\right) = 0 \\
G_1^1 &= -e^{-\nu}\left(\ddot{\omega} + \frac{3\dot{\omega}^2}{4} - \frac{\dot{\nu}\dot{\omega}}{2}\right) - 8\pi p = 0 \\
G_4^4 &= -e^{-\nu}\left(\frac{3\ddot{\omega}}{2} + \frac{3\dot{\omega}^2}{2} - \frac{3\dot{\nu}\dot{\omega}}{4}\right) + e^{-\mu}\left(\frac{3\omega^{*2}}{4} + \frac{3\omega^*\nu^*}{4}\right) = 0 \quad .
\end{aligned} \tag{2.7}$$

We see from the first of these that ρ must be positive; and from the third that p could in principle be negative, as needed in classical descriptions of particle production in quantum field theory (see, e.g., Brout, Englert and Gunzig 1978; Guth 1981; and Section 1.3). To make further progress, however, we need explicit solutions of the field equations.

A simple solution of (2.1) or (2.2) that is well known but does not depend on l has $\nu = 0$, $\omega = \log t$, $\mu = -\log t$ in (2.4), which now reads

$$dS^2 = dt^2 - t(dr^2 + r^2 d\Omega^2) - t^{-1} dl^2 \quad . \tag{2.8}$$

This has a shrinking fifth dimension, and from (2.6) or (2.7) density and pressure given by $8\pi\rho = 3/4t^2$, $8\pi p = 1/4t^2$. If these are combined to form the gravitational density $(\rho + 3p)$ and the proper radial distance $R \equiv e^{\omega/2} r$ is introduced, then the mass of a portion of the fluid is $M = 4\pi R^3(\rho + 3p)/3$. The field equations then ensure that $\ddot{R} = -M/R^2$ is obtained as usual for the law of motion. Similarly, the usual first law of thermodynamics is recovered by writing $dE + pdV = 0$ as $(\rho e^{3\omega/2})^* + p(e^{3\omega/2})^* = 0$ (E = energy, V = 3D volume; see Wesson 1992a). The equation of state of the fluid described by (2.8) is of course the $p = \rho/3$ typical of radiation.

To go beyond radiation, we use one of the classes of solutions to (2.1) or (2.2) due to Ponce de Leon (1988). With a redefinition of constants appropriate to the induced-matter theory, it has $e^\nu = l^2$, $e^\omega = t^{2/\alpha} l^{2/(1-\alpha)}$, $e^\mu = \alpha^2 (1-\alpha)^{-2} t^2$ in (2.4), which now reads

$$dS^2 = l^2 dt^2 - t^{2/\alpha} l^{2/(1-\alpha)} (dr^2 + r^2 d\Omega^2) - \alpha^2 (1-\alpha)^{-2} t^2 dl^2 \quad . \tag{2.9}$$

This has a growing fifth dimension, and density and pressure which depend on the one assignable constant α. From (2.6) or (2.7) they are

$$8\pi\rho = \frac{3}{\alpha^2 l^2 t^2} \qquad\qquad 8\pi p = \frac{(2\alpha - 3)}{\alpha^2 l^2 t^2} \quad . \tag{2.10}$$

The presence of l here may appear puzzling at first, but the coordinates are of course arbitrary and the proper time is $T \equiv lt$. (Alternatively, the presence of $x^4 = l$ depends on whether we consider the pure 4D metric or the 4D part of the 5D metric.) In terms of this, $8\pi\rho = 3/\alpha^2 T^2$ and $8\pi p = (2\alpha - 3)/\alpha^2 T^2$. For $\alpha = 3/2$, $8\pi\rho = 4/3T^2$ and $p = 0$. While for $\alpha = 2$, $8\pi\rho = 3/4T^2$ and $8\pi p = 1/4T^2$. The former is identical to the 4D Einstein-de Sitter model for the late universe with dust. The latter is identical to the 4D standard model for the early universe with radiation or highly relativistic particles. (The coincidence of the properties of matter for this model with $\alpha = 2$ and the previous model does not necessarily imply that they are the same, since similar matter can belong to different solutions even in 4D.) As before, the usual forms of the law of motion and the first law of thermodynamics are recovered, provided we use the effective gravitational density of matter defined by the combination $\rho + 3p$ (see Section 1.3; these laws are recoverable generally for metrics with form (2.4) using the proper time $T \equiv e^{\nu/2} t$ and the proper distance $R \equiv e^{\omega/2} r$). The equation of state of the fluid described by (2.9) is $p = (2\alpha/3 - 1)\rho$, and so generally describes isothermal matter.

Other properties of (2.9) were studied by Wesson (1992a, 1994), including the sizes of horizons and the nature of the extra coordinate l. We defer further discussion of this model, because here we are mainly dealing with theoretical aspects of induced-matter theory. However, we note two things. First, the solutions (2.8) and (2.9) describe in general photons with zero rest mass and particles with finite rest mass, respectively; and the fact that the former does not depend on l whereas the latter does, gives us a first inkling that l is related to mass (see later). Second, the solution (2.9) gives an excellent description of matter in the late and early universe from the big-bang perspective of physics in 4D; but it has a somewhat amazing property from the perspective of geometry in 5D. Thus consider a coordinate transformation from t, r, l to T, R, L specified by

$$
\begin{aligned}
T &= \left(\frac{\alpha}{2}\right) t^{l/\alpha} l^{l/(1-\alpha)} \left(1+\frac{r^2}{\alpha^2}\right) - \frac{\alpha}{2(1-2\alpha)} \left[t^{-l} l^{\alpha/(1-\alpha)}\right]^{(1-2\alpha)/\alpha} \\
R &= r t^{l/\alpha} l^{l/(1-\alpha)} \\
L &= \left(\frac{\alpha}{2}\right) t^{l/\alpha} l^{l/(1-\alpha)} \left(1-\frac{r^2}{\alpha^2}\right) + \frac{\alpha}{2(1-2\alpha)} \left[t^{-l} l^{\alpha/(1-\alpha)}\right]^{(1-2\alpha)/\alpha} \quad .
\end{aligned}
\tag{2.11}
$$

Then (as may be verified by computer) the Ponce de Leon metric (2.9) in standard form becomes

$$
dS^2 = dT^2 - (dR^2 + R^2 d\Omega^2) - dL^2 \quad . \tag{2.12}
$$

This means that our universe can either be viewed as a 4D spacetime curved by matter or as a 5D flat space that is empty.

2.4 The soliton case

There is a class of exact solutions of (2.1) which has been rediscovered several times during the history of Kaluza-Klein theory. The metric is static, spherically-symmetric in ordinary (3D) space, and independent of the fifth coordinate. (There are many of these solutions rather than one because Birkhoff's theorem does not apply in its conventional form in 5D.) The solutions have been interpreted as describing magnetic monopoles (Sorkin 1983), massive objects of which some called solitons have no gravitational effect (Gross and Perry 1983), and black holes (Davidson and Owen 1985). The first usage is questionable, because the 4D Schwarzschild solution is a special member of the class and gravitational in nature, and magnetic monopoles are in any case conspicuous by their absence in the real world. The last usage is misleading, because all but the Schwarzschild-like member of the class lack event horizons of the conventional sort. The middle usage can be extended, since in the induced-matter picture we will see that these solutions represent stable, extended objects (Wesson 1992b). Thus even though the word is over-worked in physics, we regard these 5D 1-body solutions as representing objects called solitons.

A particularly simple member of the soliton class was rediscovered by Chatterjee (1990). It is instructive to start with this, because it has been analyzed in standard or Schwarzschild-like coordinates (as opposed to the isotropic coordinates used below). Thus the 5D metric has an interval given by

$$
\begin{aligned}
dS^2 = \ & \left[1 - \frac{2\sqrt{A}}{\sqrt{r^2+A}+\sqrt{A}}\right] dt^2 - \frac{dr^2}{1+A/r^2} - r^2 d\Omega^2 \\
& - \left[1 - \frac{2\sqrt{A}}{\sqrt{r^2+A}+\sqrt{A}}\right]^{-1} dl^2 \quad .
\end{aligned}
\tag{2.13}
$$

Here the constant A would normally be identified in gravitational applications via the $r \to \infty$ limit as $\sqrt{A} = M_*$, the mass of an object like a star at the centre of ordinary space. However, as mentioned above, it cannot be assumed that (2.13) is a black hole. Indeed, the first metric coefficient in (2.13) goes to zero only for r tending to zero. In other words, the event horizon in the coordinates of (2.13) shrinks to a point at the centre of ordinary space. (This is not altered by a Killing-vector prescription for horizons and different sets of coordinates: see Wesson and Ponce de Leon 1984.) To see what (2.13) actually represents, let us use the induced-matter approach. It is helpful to consider a metric we will come back to later, namely

$$dS^2 = e^{\nu}dt^2 - e^{\lambda}dr^2 + R^2 d\Omega^2 - e^{\mu}dl^2 \quad . \tag{2.14}$$

This includes (2.13) if we assume that the metric coefficients ν, λ, R, μ depend only on the radius and not on the time or extra coordinate. Then following the same procedure as in the preceding section, we obtain the components of the induced 4D energy-momentum tensor:

$$\begin{aligned}
8\pi T_0^0 &= e^{-\lambda}\left(\frac{1}{2}\mu'' + \frac{1}{4}\mu'^2 + \frac{R'\mu'}{R} - \frac{1}{4}\lambda'\mu'\right) \\
8\pi T_1^1 &= e^{-\lambda}\left(\frac{R'\mu'}{R} + \frac{1}{4}\nu'\mu'\right) \\
8\pi T_2^2 &= e^{-\lambda}\left(\frac{1}{2}\mu'' + \frac{1}{4}\mu'^2 + \frac{R'\mu'}{2R} + \frac{1}{4}\nu'\mu' - \frac{1}{4}\lambda'\mu'\right) \quad .
\end{aligned} \tag{2.15}$$

Here a prime denotes the partial derivative with respect to the radius. Other components are zero, and of course $T_3^3 = T_2^2$ because of spherical symmetry. The components (2.15) which define the properties of matter depend on derivatives of μ, that is upon the geometry of the fifth dimension. However, matter and geometry are unified via the field equations $G_B^A = 0$ of (2.2), and we can use these to rewrite (2.15) in a more algebraically convenient form:

$$\begin{aligned}
8\pi T_0^0 &= \frac{1}{R^2} - e^{-\lambda}\left(\frac{2R''}{R} + \frac{R'^2}{R^2} - \frac{R'\lambda'}{R}\right) \\
8\pi T_1^1 &= \frac{1}{R^2} - e^{-\lambda}\left(\frac{R'^2}{R^2} + \frac{R'\nu'}{R}\right) \\
8\pi T_2^2 &= -\frac{1}{4}e^{-\lambda}\left(2\nu'' + \nu'^2 + \frac{4R''}{R} - \frac{2R'\lambda'}{R} + \frac{2R'\nu'}{R} - \nu'\lambda'\right) \quad .
\end{aligned} \tag{2.16}$$

Substitutinginto these equations for the Chattejee solution (2.13) gives

$$\begin{aligned}
8\pi T_0^0 &= \frac{A}{r^4} \\
8\pi T_1^1 &= -\frac{2\sqrt{A}}{r^2\sqrt{r^2+A}} - \frac{2A\sqrt{A}}{r^4\sqrt{r^2+A}} - \frac{A}{r^4} \\
8\pi T_2^2 &= \frac{\sqrt{A}}{r^2\sqrt{r^2+A}} + \frac{A\sqrt{A}}{r^4\sqrt{r^2+A}} \quad .
\end{aligned} \tag{2.17}$$

These components obey the equation of state

$$T_0^0 + T_1^1 + T_2^2 + T_3^3 = 0 \quad , \tag{2.18}$$

which is radiation-like. The matter described by (2.17) has a gravitational mass that can be evaluated using the standard 4D expression

$$M_g(r) \equiv \int (T_0^0 - T_1^1 - T_2^2 - T_3^3)\sqrt{-g_4}dV_3 \quad , \tag{2.19}$$

where g_4 is the determinant of the 4-metric and dV_3, is a 3D volume element. Using (2.14) and (2.16) this gives

$$M_g(r) = \frac{1}{2}\int\left(\nu'' + \frac{1}{2}\nu'^2 + \frac{2R'\nu'}{R} - \frac{1}{2}\nu'\lambda'\right)e^{(\nu-\lambda)/2}R^2 dr \quad . \tag{2.20}$$

This is most conveniently evaluated in the coordinates of the Chatterjee solution in the form (2.13). Thus putting $R \to r$ and integrating gives

$$M_g(r) = \frac{1}{2}r^2 e^{(\nu-\lambda)/2}\nu' \quad , \tag{2.21}$$

which with the coefficients of (2.13) is

$$M_g(r) = \sqrt{A}\left[\frac{\sqrt{r^2+A}-\sqrt{A}}{\sqrt{r^2+A}+\sqrt{A}}\right]^{1/2} \quad . \tag{2.22}$$

We see that $M_g(\infty) = \sqrt{A}$, agreeing with the usual metric-based definition of the mass as noted above. However, we also see that $M_g(0) = 0$, meaning that the gravitational mass goes to zero at the centre. In summary, the Chatterjee soliton (2.13) is a ball of radiation-like matter whose density and pressure fall off very rapidly away from the centre, and whose integrated mass agrees with the conventionai definition only at infinity.

The above concerned a special case of a broad class of 5D solutions which has been widely studied in forms due to Gross and Perry (1983) and Davidson and Owen (1985). These authors use different terminologies, particularly for two dimensionless constants which enter the solutions. The former use α, β and the latter use κ, ε where the two are related by $\kappa = -1/\beta$, $\varepsilon = -\beta/\alpha$ We adopt the latter notation, as it is more suited to the induced-matter approach. In it, positive effective density of matter requires $\kappa > 0$, and positive gravitational mass as measured at spatial infinity requires $\varepsilon\kappa > 0$ (see below). Thus, physicality requires that both κ and ε be positive. In terms of these constants the Chatterjee solution we have looked at already has just $\kappa = 1$, $\varepsilon = 1$. And the Schwanschild solution we will look at below has $\varepsilon \to 0$, $\kappa \to \infty$, $\varepsilon\kappa \to 1$. We now proceed to consider the general class.

This has usually been discussed with the metric in spatially isotropic form, which we write as

$$dS^2 = e^\nu dt^2 - e^\lambda(dr^2 + r^2 d\Omega^2) - e^\mu dl^2 \quad . \tag{2.23}$$

Then solutions of the apparently empty 5D field equations (2.1) or (2.2) are given by

$$\begin{aligned}
e^{\nu/2} &= \left(\frac{ar-1}{ar+1}\right)^{\varepsilon\kappa} \\
e^{\lambda/2} &= \frac{(ar-1)(ar+1)}{a^2r^2}\left(\frac{ar+1}{ar-1}\right)^{\ell(\kappa-1)} \\
e^{\mu/2} &= \left(\frac{ar+1}{ar-1}\right)^{\varepsilon} \quad .
\end{aligned} \tag{2.24}$$

Here a is a dimensional constant to do with the source, and the two dimensionless constants are related by a consistency relation derived from the field equations:

$$\varepsilon^2(\kappa^2 - \kappa + 1) = 1 \quad . \tag{2.25}$$

This means that the class is a 2-parameter one, depending on a and one or the other of ε, κ. Also, we noted above that physicality requires that both κ and ε be positive. Now, the surface area of

2-shells around the centre of the 3-geometry varies as $(ar-1)^{1-\varepsilon(\kappa-1)}$, and will shrink to zero at $r = 1/a$ provided $1-\varepsilon(\kappa-1) > 0$. This combined with (2.25) means $\kappa > 0$. That is, the centre of the 3-geometry is at $r = l/a$ for physical choices of the parameters (see Billyard, Wesson and Kalligas 1995 for a more extensive discussion). Also, $e^{\nu} \to 0$ for $r \to 1/a$ for ε, $\kappa > 0$. So as for the Chatterjee case above, the event horizon for the general class shrinks to a point at the centre of ordinary space.

The properties of the induced matter associated with (2.24) can be worked out following the same procedure as before. The components of the induced 4D energy-momentum tensor are:

$$\begin{aligned}
8\pi T^0_0 &= -e^{\lambda}\left(\lambda'' + \frac{1}{4}\lambda'^2 + \frac{2\lambda'}{r}\right) \\
8\pi T^1_1 &= -e^{\lambda}\left(\frac{1}{4}\lambda'^2 + \frac{1}{2}\nu'\lambda' + \frac{\nu'}{r} + \frac{\lambda'}{r}\right) \\
8\pi T^2_2 &= -e^{\lambda}\left(\nu'' + \lambda'' + \frac{1}{2}\nu'^2 + \frac{\nu'}{r} + \frac{\lambda'}{r}\right) \quad .
\end{aligned} \tag{2.26}$$

Substituting into these equations for the solutions in the form (2.24) and doing some tedious algebra gives

$$\begin{aligned}
8\pi T^0_0 &= \frac{4\varepsilon^2\kappa a^6 r^4}{(ar-1)^4(ar+1)^4}\left(\frac{ar-1}{ar+1}\right)^{2\varepsilon(\kappa-1)} \\
8\pi T^1_1 &= \frac{4\varepsilon a^5 r^3}{(ar-1)^3(ar+1)^3}\left(\frac{ar-1}{ar+1}\right)^{2\varepsilon(\kappa-1)} \\
&\quad -\frac{4\varepsilon a^6 r^4(2\varepsilon + 2ar - \varepsilon\kappa)}{(ar-1)^4(ar+1)^4}\left(\frac{ar-1}{ar+1}\right)^{2\varepsilon(\kappa-1)} \\
8\pi T^2_2 &= -\frac{2\varepsilon a^5 r^3}{(ar-1)^3(ar+1)^3}\left(\frac{ar-1}{ar+1}\right)^{2\varepsilon(\kappa-1)} \\
&\quad -\frac{4\varepsilon a^6 r^4(\varepsilon\kappa - \varepsilon + ar)}{(ar-1)^4(ar+1)^4}\left(\frac{ar-1}{ar+1}\right)^{2\varepsilon(\kappa-1)} \quad .
\end{aligned} \tag{2.27}$$

These components obey the same equation of state as before, namely $(T^0_0 + T^1_1 + T^2_2 + T^3_3) = 0$. If we average over the 3 spatial directions, this is equivalent to saying that the equation of state is $\overline{p} = \rho/3$. Also as before, we can calculate the standard 4D gravitational mass of a part of the fluid by using (2.19). This with (2.23) gives

$$M_g(r) = 4\pi \int (T^0_0 - T^1_1 - T^2_2 - T^3_3) e^{(\nu+3\lambda)/2} r^2 dr \quad , \tag{2.28}$$

which with (2.26) is

$$\begin{aligned}
M_g(r) &= \frac{1}{2}\int \left(\nu'' + \frac{1}{2}\nu'^2 + \frac{2\nu'}{r} + \frac{1}{2}\nu'\lambda'\right) e^{(\nu+\lambda)/2} r^2 dr \\
&= \frac{1}{2} r^2 e^{(\nu-\lambda)/2}\nu' \quad .
\end{aligned} \tag{2.29}$$

Then with the coefficients of (2.24) we obtain

$$M_g(r) = \frac{2\varepsilon\kappa}{a}\left(\frac{ar-1}{ar+1}\right)^{\varepsilon} \quad . \tag{2.30}$$

This is the gravitational mass of a soliton as a function of (isotropic) radius r, and to be positive as measured at infinity requires that $\varepsilon\kappa > 0$. Since positive density requires that $\kappa > 0$ by (2.27) we see that we need both $\kappa > 0$ and $\varepsilon > 0$, as we stated above. Then (2.30) shows that $M_g(r = 1/a) = 0$, meaning again that the gravitational mass goes to zero at the centre. However, the mass as measured at spatial infinity is now $2\varepsilon\kappa/a$ and not just $2/a = M_*$ as it was for the Chatterjee case.

This is interesting, and should be compared to what we obtain if we substitute parameters corresponding to the Schwarzschild case, namely $\varepsilon \to 0$, $\kappa \to \infty$, $\varepsilon\kappa \to 1$. Then (2.30) gives $M_g(r) =$ constant $= 2/a = M_*$. And the metric (2.23), (2.24) becomes

$$dS^2 = \left(\frac{1 - M_*/2r}{1 + M_*/2r}\right)^2 dt^2 - \left(1 + \frac{M_*}{2r}\right)^4 (dr^2 + r^2 d\Omega^2) - dl^2 \quad . \tag{2.31}$$

This is just the Schwarzschild solution (in isotropic coordinates) plus a flat and therefore physically innocuous extra dimension. In other words, if we use the conventionally defined 4D gravitational mass as a diagnostic for 5D solitons, we recover the usual 4D Schwarzschild mass exactly.

We have carried out a numerical investigation of preceding relations to clarify the status of the Schwarzschild solution (Wesson and Ponce de Leon 1994). The problem is that if we *set* $\varepsilon = 0$ and $\varepsilon\kappa = 1$ then (2.30) gives $M_g(r) = 2/a$ for all r; but if we keep ε *small* and let $r \to 1/a$, then (2.30) gives $M_g(r = 1/a) = 0$ irrespective of ε. Clearly the limit by which one is supposed to recover the Schwanschild solution from the soliton solutions is ambiguous. However, our numerical results show that, from the viewpoint of perturbation analysis at least, the Schwarzschild case is just a highly compressed soliton. We have also looked at other definitions for the mass of a soliton, including the so-called proper mass (which depends on an integral involving only and is badly defined at the centre) and the ADM mass (which depends on a product of field strength and area and is well defined at the centre). To clarify what happens near the centre of a soliton, we have also calculated the geometric scalars for metric (2.23), (2.24). The relevant 5D invariant is the Kretschmann scalar $K \equiv R_{ABCD}R^{ABCD}$, which we have evaluated algebraically and checked by computer. It is

$$\begin{aligned} K &= \frac{192a^{10}r^6}{(a^2r^2-1)^8}\left(\frac{ar-1}{ar+1}\right)^{4\varepsilon(\kappa-1)} \{1 - 2\varepsilon(\kappa-1)(2+\varepsilon^2\kappa)ar \\ &\quad +2(3-\varepsilon^4\kappa^2)a^2r^2 - 2\varepsilon(\kappa-1)(2+\varepsilon^2\kappa)a^3r^3 + a^4r^4\} \quad . \end{aligned} \tag{2.32}$$

Taking into account the constraint (2.25), this may be found to diverge for $\kappa > 0$ at $r = 1/a$, showing that there is a geometric singularity at the centre as we have defined the latter. [Note that if we let $\varepsilon \to 0$, $\kappa \to \infty$, $\varepsilon\kappa \to 1$ then (2.32) gives $K = 192a^{10}r^6/(ar+1)^{12}$ in isotropic coordinates, or $K = 48M_*^2/(r')^6$ in curvature coordinates where $r' = r(1 + 1/ar)^2$ and $a = 2/M_*$. This result agrees formally with the one from Einstein theory, but from our viewpoint no longer has much significance since the point $r' = 0$ or $r = -1/a$ is not part of the manifold, which ends at $r = 1/a$ or $r' = 2M_*$.] The relevant 4D invariant is $C \equiv R_{\alpha\beta}R^{\alpha\beta}$, which is most easily evaluated algebraically using the field equations. The latter are (2.1) or $R_{AB} = 0$, but if there is no dependency on the extra coordinate read just $R_{\alpha\beta} = 8\pi T_{\alpha\beta}$ because the 4D Ricci scalar R is zero. Then $C = 64\pi^2[(T_0^0)^2 + (T_1^1)^2 + 2(T_2^2)^2]$, and can be evaluated using the components of the energy-momentum tensor (2.27). It is

$$C = \frac{8\varepsilon^2a^{10}r^6}{(a^2r^2-1)^8}\left(\frac{ar-1}{ar+1}\right)^{2\varepsilon(\kappa-1)} \{3 + 4\varepsilon(3-2\kappa)ar$$

$$+2(3+6\varepsilon^2+4\varepsilon^2\kappa^2-8\varepsilon^2\kappa)a^2r^2+12\varepsilon a^3r^3+3a^4r^4\} \quad . \tag{2.33}$$

This also diverges at $r = 1/a$, confirming that there is a singularity in the geometry at the centre. In conjunction with the fact that most of them do not have event horizons of the standard sort, this means that technically solitons should be classified as naked singularities.

Whether or not we can see to the centre of a soliton, practically, is a different question. What was a point mass in 4D general relativity has become a finite object in 5D induced-matter theory. The fluid is 'hot', with anisotropic pressure and density that falls off rapidly away from the centre (for large distances it goes as M_*^2/r^4 where M_* is the mass as measured at spatial infinity). The Schwarzschild solution is somewhat anomalous, but can be regarded as a soliton where matter is so concentrated towards the centre as to leave most of space empty. In short, solitons are 'holes' in the geometry surrounded by induced matter.

2.5 The case of neutral matter

Kaluza-Klein theory in 5D has traditionally identified the $g_{4\alpha}$ components of the metric tensor with the potentials A_α of classical electromagnetism (see Section 1.5). These set to zero therefore give in some sense a description of neutral matter. However, any fully covariant 5D theory, such as induced-matter theory, has 5 coordinate degrees of freedom, which used judiciously can lead to considerable algebraic simplification without loss of generality. Therefore, a natural case to study is specified by $g_{4\alpha} = 0$, $g_{44} \neq 0$. This removes the explicit electromagnetic potentials and leaves one coordinate degree of freedom over to be used appropriately (e.g., to simplify the equation of motion of a particle). This choice of coordinates or choice of gauge involves $g_{\alpha\beta} = g_{\alpha\beta}(x^A)$, $g_{44} = g_{44}(x^A)$ and so is not restricted by the cylinder condition of old Kaluza-Klein theory. It admits fluids consisting of particles with finite or zero rest mass, and thus includes the cases we have studied in Sections 2.3 and 2.4. In the present section, we follow Wesson and Ponce de Leon (1992). Our aims are to give a reasonably self-contained account of the matter gauge and to lay the foundation for later applications.

We have a 5D interval $dS^2 = g_{AB}dx^A dx^B$ where

$$\begin{aligned} g_{\alpha\beta} &= g_{\alpha\beta}(x^A) & \qquad g_{4\alpha} &= 0 \\ g_{44} &\equiv \varepsilon\Phi^2(x^A) & \qquad g^{44} &= \frac{1}{g_{44}} = \frac{\varepsilon}{\Phi^2} \quad . \end{aligned} \tag{2.34}$$

Here $\varepsilon^2 = 1$ and the signature of the scalar part of the metric is left general. (We will see later that there are well-behaved classical solutions of the field equations with $\varepsilon = +1$ as well as the often-assumed $\varepsilon = -1$, and the freedom to choose this may also help with the Euclidean approach to quantum gravity.) The 5D Ricci tensor in terms of the 5D Christoffel symbols is given by

$$R_{AB} = (\Gamma^C_{AB})_{,C} - (\Gamma^C_{AC})_{,B} + \Gamma^C_{AB}\Gamma^D_{CD} - \Gamma^C_{AD}\Gamma^D_{BC} \quad . \tag{2.35}$$

Here a comma denotes the partial derivative, and below we will use a semicolon to denote the ordinary (4D) covariant derivative. Putting $A \to \alpha$, $B \to \beta$ in (2.35) gives us the 4D part of the 5D quantity. Expanding some summed terms on the r.h.s. by letting $C \to \lambda, 4$ etc. and rearranging gives

$$\begin{aligned} \hat{R}_{\alpha\beta} &= (\Gamma^\lambda_{\alpha\beta})_{,\lambda} + (\Gamma^4_{\alpha\beta})_{,4} - (\Gamma^\lambda_{\alpha\lambda})_{,\beta} - (\Gamma^4_{\alpha4})_{,\beta} + \Gamma^\lambda_{\alpha\beta}\Gamma^\mu_{\lambda\mu} \\ &+ \Gamma^\lambda_{\alpha\beta}\Gamma^4_{\lambda4} + \Gamma^4_{\alpha\beta}\Gamma^D_{4D} - \Gamma^\mu_{\alpha\lambda}\Gamma^\lambda_{\beta\mu} - \Gamma^4_{\alpha\lambda}\Gamma^\lambda_{\beta4} - \Gamma^D_{\alpha4}\Gamma^4_{\beta D} \quad . \end{aligned} \tag{2.36}$$

Part of this is the conventional Ricci tensor that only depends on indices 0123, so

$$\begin{aligned}\hat{R}_{\alpha\beta} &= R_{\alpha\beta} + (\Gamma^4_{\alpha\beta})_{,4} - (\Gamma^4_{\alpha 4})_{,\beta} + \Gamma^\lambda_{\alpha\beta}\Gamma^4_{\lambda 4} \\ &+ \Gamma^4_{\alpha\beta}\Gamma^D_{4D} - \Gamma^4_{\alpha\lambda}\Gamma^\lambda_{\beta 4} - \Gamma^D_{\alpha 4}\Gamma^4_{\beta D} \quad .\end{aligned} \tag{2.37}$$

To evaluate this we need the Christoffel symbols.

These can be tabulated here in appropriate groups:

$$\begin{aligned}\Gamma^4_{\alpha\beta} &= -\frac{g^{44}g^*_{\alpha\beta}}{2} & \Gamma^4_{\alpha 4} &= \frac{g^{44}g_{44,\alpha}}{2} \\ \Gamma^D_{4D} &= \frac{g^{DC}g^*_{DC}}{2} & \Gamma^\lambda_{\beta 4} &= \frac{g^{\lambda C}g^*_{\beta C}}{2} \\ \Gamma^D_{\alpha 4} &= \frac{g^{D4}g_{44,\alpha}}{2} + \frac{g^{D\gamma}g^*_{\alpha\gamma}}{2} & \Gamma^4_{\beta D} &= \frac{g^{44}g_{D4,\beta}}{2} - \frac{g^{44}g_{\beta D,4}}{2} \quad .\end{aligned} \tag{2.38}$$

$$\begin{aligned}\Gamma^\lambda_{44} &= -\frac{g^{\lambda\beta}g_{44,\beta}}{2} & \Gamma^\lambda_{4\lambda} &= \frac{g^{\lambda\beta}g^*_{\lambda\beta}}{2} \\ \Gamma^\mu_{\lambda\mu} &= \frac{g^{\mu\beta}g_{\mu\beta,\lambda}}{2} & \Gamma^4_{44} &= \frac{g^{44}g^*_{44}}{2} \\ \Gamma^\lambda_{4\mu} &= \frac{g^{\lambda\beta}g^*_{\mu\beta}}{2} & \Gamma^4_{4\mu} &= \frac{g^{44}g_{44,\mu}}{2} \quad .\end{aligned} \tag{2.39}$$

$$\begin{aligned}\Gamma^\lambda_{4\alpha} &= \frac{g^{\lambda\mu}g^*_{\alpha\mu}}{2} & \Gamma^4_{4\alpha} &= \frac{g^{44}g_{44,\alpha}}{2} \\ \Gamma^\lambda_{4\lambda} &= \frac{g^{\lambda\beta}g^*_{\lambda\beta}}{2} & \Gamma^4_{44} &= \frac{g^{44}g_{44,4}}{2} \\ \Gamma^\mu_{\lambda\mu} &= \frac{g^{\mu\nu}g_{\mu\nu,\lambda}}{2} & \Gamma^\mu_{\alpha\lambda} &= \frac{g^{\mu\sigma}}{2}(g_{\lambda\sigma,\alpha} + g_{\sigma\alpha,\lambda} - g_{\alpha\lambda,\sigma}) \\ \Gamma^\lambda_{44} &= -\frac{g^{\alpha\lambda}g_{44,\alpha}}{2} & \Gamma^4_{\alpha\lambda} &= -\frac{g^{44}g^*_{\alpha\lambda}}{2} \quad .\end{aligned} \tag{2.40}$$

We will use these respectively to evaluate (2.37) above and (2.44), (2.54) below.

Thus substituting into and expanding some terms in (2.37) gives

$$\begin{aligned}\hat{R}_{\alpha\beta} &= R_{\alpha\beta} - \frac{g^{*44}g^*_{\alpha\beta}}{2} - \frac{g^{44}g^{**}_{\alpha\beta}}{2} - \frac{g^{44}_{,\beta}g_{44,\alpha}}{2} - \frac{g^{44}g_{44,\alpha\beta}}{2} \\ &+ \frac{g^{44}g_{44,\lambda}\Gamma^\lambda_{\alpha\beta}}{2} - \frac{g^{\mu\nu}g^*_{\mu\nu}g^{44}g^*_{\alpha\beta}}{4} - \frac{(g^{44})^2 g^*_{\alpha\beta}g^*_{44}}{4} \\ &+ \frac{g^{\lambda\mu}g^{44}g^*_{\alpha\lambda}g^*_{\beta\mu}}{2} - \frac{(g^{44})^2 g_{44,\alpha}g_{44,\beta}}{4} \quad .\end{aligned} \tag{2.41}$$

Some of the terms here may be rewritten using

$$
\begin{aligned}
-\frac{g^{44}_{,\beta}g_{44,\alpha}}{2} &- \frac{g^{44}g_{44,\alpha\beta}}{2} + \frac{g^{44}g_{44,\lambda}\Gamma^{\lambda}_{\alpha\beta}}{2} - \frac{(g^{44})^2 g_{44,\alpha}g_{44,\beta}}{4} \\
&= -\frac{1}{\Phi}(\Phi_{\alpha;\beta} - \Phi_{\lambda}\Gamma^{\lambda}_{\alpha\beta}) \equiv -\frac{\Phi_{\alpha;\beta}}{\Phi} \quad ,
\end{aligned} \tag{2.42}
$$

Where $\Phi_\alpha \equiv \Phi_{,\alpha}$. Then (2.41) gives

$$
\hat{R}_{\alpha\beta} = R_{\alpha\beta} - \frac{\Phi_{\alpha;\beta}}{\Phi} + \frac{\varepsilon}{2\phi^2}\left(\frac{\Phi^* g^*_{\alpha\beta}}{\Phi} - g^{**}_{\alpha\beta} + g^{\lambda\mu}g^*_{\alpha\lambda}g^*_{\beta\mu} - \frac{g^{\mu\nu}g^*_{\mu\nu}g^*_{\alpha\beta}}{2}\right) \quad . \tag{2.43}
$$

We will use this below when we consider the field equations.

Returning to (2.35), we put $A = 4$, $B = 4$ and expand with $C \to \lambda$, 4 etc. to obtain

$$
R_{44} = (\Gamma^{\lambda}_{44})_{,\lambda} - (\Gamma^{\lambda}_{4\lambda})_{,4} + \Gamma^{\lambda}_{44}\Gamma^{\mu}_{\lambda\mu} + \Gamma^{4}_{44}\Gamma^{\mu}_{4\mu} - \Gamma^{\lambda}_{4\mu}\Gamma^{\mu}_{4\lambda} - \Gamma^{4}_{4\mu}\Gamma^{\mu}_{44} \quad . \tag{2.44}
$$

The Christoffel symbols here are tabulated in (2.39), and cause (2.44) to become

$$
\begin{aligned}
R_{44} = &- \frac{g^{\lambda\beta}_{,\lambda}g_{44,\beta}}{2} - \frac{g^{\lambda\beta}g_{44,\beta\lambda}}{2} - \frac{g^{*\lambda\beta}g^*_{\lambda\beta}}{2} - \frac{g^{\lambda\beta}g^{**}_{\lambda\beta}}{2} \\
&- \frac{g^{\lambda\beta}g_{44,\beta}g^{\mu\sigma}g_{\mu\sigma,\lambda}}{4} + \frac{g^{44}g^*_{44}g^{\lambda\beta}g^*_{\lambda\beta}}{4} \\
&- \frac{g^{\mu\beta}g^*_{\lambda\beta}g^{\lambda\sigma}g^*_{\mu\sigma}}{4} + \frac{g^{44}g_{44,\lambda}g^{\lambda\beta}g_{44,\beta}}{4} \quad .
\end{aligned} \tag{2.45}
$$

Some of the terms here may be rewritten using

$$
\begin{aligned}
-\frac{g^{\lambda\beta}g_{44,\beta}}{2} &- \frac{g^{\lambda\beta}g_{44,\beta\lambda}}{2} - \frac{g^{\lambda\beta}g_{44,\beta}g^{\mu\sigma}g_{\mu\sigma,\lambda}}{4} + \frac{g^{44}g_{44,\lambda}g^{\lambda\beta}g_{44,\beta}}{4} \\
&= -\varepsilon\Phi\left(g^{\lambda\beta}_{,\lambda}\Phi_{\beta} + g^{\lambda\beta}\Phi_{\beta,\lambda} + \frac{g^{\lambda\beta}g^{\mu\sigma}g_{\mu\sigma,\lambda}\Phi_{\beta}}{2}\right) \\
&= -\varepsilon\Phi g^{\mu\nu}\Phi_{\mu;\nu} \quad .
\end{aligned} \tag{2.46}
$$

Here we have obtained the last line by noting that $\Phi_{\beta;\lambda} = \Phi_{\beta,\lambda} - \Gamma^{\sigma}_{\beta\lambda}\Phi_{\sigma}$ implies

$$
g^{\beta\lambda}\Phi_{\beta,\lambda} + \frac{g^{\beta\lambda}g^{\sigma\mu}g_{\beta\lambda,\mu}\Phi_{\sigma}}{2} = g^{\mu\nu}\Phi_{\mu;\nu} + \frac{g^{\beta\lambda}g^{\sigma\mu}g_{\mu\beta,\lambda}\Phi_{\sigma}}{2} + \frac{g^{\beta\lambda}g^{\sigma\mu}g_{\mu\lambda,\beta}\Phi_{\sigma}}{2} \quad , \tag{2.47}
$$

and that $(\sigma^{\mu}_{\nu})_{,\mu} = 0$ implies $(g^{\lambda\sigma}_{,\lambda} + g^{\beta\lambda}g^{\sigma\mu}g_{\mu\beta,\lambda})\Phi_{\sigma} = 0$. Putting (2.46) in (2.45) gives lastly

$$
R_{44} = -\varepsilon\Phi\Box\Phi - \frac{g^{*\lambda\beta}g^*_{\lambda\beta}}{2} - \frac{g^{\lambda\beta}g^{**}_{\lambda\beta}}{2} + \frac{\Phi^* g^{\lambda\beta}g^*_{\lambda\beta}}{2\Phi} - \frac{g^{\mu\beta}g^{\lambda\sigma}g^*_{\lambda\beta}g^*\mu\sigma}{4} \quad , \tag{2.48}
$$

where $\Box\Phi \equiv g^{\mu\nu}\Phi_{\mu;\nu}$ defines the 4D curved-space box operator. Equations (2.43) and (2.48) can be used with the 5D field equations (2.1) which we repeat here:

$$
R_{AB} = 0 \quad . \tag{2.49}
$$

Then $\hat{R}_{\alpha\beta} = 0$ in (2.43) gives

$$R_{\alpha\beta} = \frac{\Phi_{\alpha;\beta}}{\Phi} - \frac{\varepsilon}{2\Phi^2}\left(\frac{\Phi^* g^*_{\alpha\beta}}{\Phi} - g^{**}_{\alpha\beta} + g^{\lambda\mu} g^*_{\alpha\lambda} g^*_{\beta\mu} - \frac{g^{\mu\nu} g^*_{\mu\nu} g^*_{\alpha\beta}}{2}\right) \quad . \tag{2.50}$$

And $R_{44} = 0$ in (2.48) gives

$$\varepsilon\Phi\Box\Phi = -\frac{g^{*\lambda\beta} g^*_{\lambda\beta}}{4} - \frac{g^{\lambda\beta} g^{**}_{\lambda\beta}}{2} + \frac{\Phi^* g^{\lambda\beta} g^*_{\lambda\beta}}{2\Phi} \quad , \tag{2.51}$$

where we have noted that $(\delta^\mu_\nu)_{,4} = 0$ implies $g^{\mu\beta} g^{\lambda\sigma} g^*_{\lambda\beta} g^*_{\mu\sigma} + g^{*\mu\sigma} g^*_{\mu\sigma} = 0$. From (2.50) we can form the 4D Ricci curvature scalar $R = g^{\alpha\beta} R_{\alpha\beta}$. Eliminating the covariant derivative using (2.51), and again using $(\delta^\mu_\nu)_{,4} = 0$ to eliminate some terms, gives

$$R = \frac{\varepsilon}{4\Phi^2}\left[g^{*\mu\nu} g^*_{\mu\nu} + \left(g^{\mu\nu} g^*_{\mu\nu}\right)^2\right] \quad . \tag{2.52}$$

With (2.50) and (2.52), we are now in a position to define if we wish an energy-momentum tensor in 4D via $8\pi T_{\alpha\beta} \equiv R_{\alpha\beta} - R g_{\alpha\beta}/2$. It is

$$\begin{aligned} 8\pi T_{\alpha\beta} = \; & \frac{\Phi_{\alpha;\beta}}{\Phi} - \frac{\varepsilon}{2\Phi^2}\left\{\frac{\Phi^* g^*_{\alpha\beta}}{\Phi} - g^{**}_{\alpha\beta} + g^{\lambda\mu} g^*_{\alpha\lambda} g^*_{\beta\mu} - \frac{g^{\mu\nu} g^*_{\mu\nu} g^*_{\alpha\beta}}{2}\right. \\ & \left. + \frac{g_{\alpha\beta}}{4}\left[g^{*\mu\nu} g^*_{\mu\nu} + (g^{\mu\nu} g^*_{\mu\nu})^2\right]\right\} \quad . \end{aligned} \tag{2.53}$$

Provided we use this energy-momentum tensor, Einstein's 4D field equations (2.3) or $G_{\alpha\beta} = 8\pi T_{\alpha\beta}$ will of course be satisfied.

The mathematical expression (2.53) has good properties. It is a symmetric tensor that has a part which depends on derivatives of Φ with respect to the usual coordinates x^{0123}, and a part which depends on derivatives of other metric coefficients with respect to the extra coordinate x^4. [The first term in (2.53) is implicitly symmetric because it depends on the second partial derivative, while the other terms are explicitly symmetric.] It is also compatible with what is known about the recovery of 4D properties of matter from apparently empty 5D solutions of Kaluza-Klein theory. Thus the cosmological case studied in Section 2.3 agrees with (2.53) and has matter which owes its characteristics largely to the x^4-dependency of $g_{\alpha\beta}$ in that relation. While the soliton case studied in Section 2.4 agrees with (2.53) and has matter which depends on the first or scalar term in that relation. With (2.53) and preceding relations, the case where there is no dependency on x^4 becomes transparent. Then (2.51) becomes the scalar wave equation for the extra part of the metric ($g^{\mu\nu}\Phi_{\mu;\nu} = 0$ with $g_{44} = \varepsilon\Phi^2$). And (2.53) gives $T = T_{\alpha\beta} g^{\alpha\beta} = 0$, which implies a radiation-like equation of state. However, in general there must be x^4-dependence if we are to recover more complex equations of state from solutions of $R_{AB} = 0$.

These field equations have 4 other components we have not so far considered, namely $R_{4\alpha} = 0$. This relation by (2.35) expanded is

$$\begin{aligned} R_{4\alpha} \; &= \; (\Gamma^\lambda_{4\alpha})_{,\lambda} + (\Gamma^4_{4\alpha})_{,4} - (\Gamma^\lambda_{4\lambda})_{,\alpha} - (\Gamma^4_{44})_{,\alpha} \\ &+ \; \Gamma^\lambda_{4\lambda}\Gamma^A_{\lambda A} + \Gamma^4_{4\alpha}\Gamma^A_{4A} - \Gamma^A_{4\mu}\Gamma^\mu_{\alpha A} - \Gamma^D_{44}\Gamma^4_{\alpha D} \quad . \end{aligned} \tag{2.54}$$

The Christoffel symbols here are tabulated in (2.40) and cause (2.54) to become

$$
\begin{aligned}
R_{4\alpha} = \quad & \frac{g^{44}g^{\lambda\beta}}{4}\left(g^*_{\lambda\beta}g_{44,\alpha} - g_{44,\beta}g^*_{\alpha\lambda}\right) + \frac{g^{\lambda\mu}_{,\lambda}\, g^*_{\mu\alpha}}{2} \\
+ \quad & \frac{g^{\lambda\mu}g^*_{\mu\alpha,\lambda}}{2} - \frac{g^{\lambda\beta}_{,\alpha}\, g^*_{\lambda\beta}}{2} - \frac{g^{\lambda\beta}g^*_{\lambda\beta,\alpha}}{2} \\
+ \quad & \frac{g^{\lambda\sigma}g^{\mu\beta}g^*_{\sigma\alpha}g_{\mu\beta,\lambda}}{4} + \frac{g^{*\mu\beta}g_{\mu\beta,\alpha}}{4} \quad . \qquad (2.55)
\end{aligned}
$$

Here we have done some algebra using $(g_{44}g^{44})_{,\alpha \ and \ 4} = 0$ or $g^{*44}g_{44,\alpha} - g^{44}_{,\alpha}g^*_{44} = 0$, and $(\delta^{\beta}_{\lambda})_{,4} = 0$ or $g^{\lambda\nu}g_{\mu\nu}g^{\mu\sigma} + g^{*\lambda\sigma} = 0$. [We also note in passing that one can use $(\Gamma^4_{4\alpha})_{,4} = (\Gamma^4_{44})_{,\alpha}$ in (2.54) and obtain an alternative form of (2.55) with the last term replaced by $g^{\mu\beta}_{,\alpha}g^*_{\mu\beta}/4$.] While (2.55) may be useful in other computations, it is helpful for our purpose here to rewrite it as

$$
\begin{aligned}
R_{4\alpha} \quad = \quad & \frac{\partial}{\partial x^{\beta}}\left(\frac{g^{\beta\lambda}g^*_{\lambda\alpha}}{2}\right) - \frac{\partial}{\partial x^{\alpha}}\left(\frac{g^{\mu\nu}g^*_{\mu\nu}}{2}\right) + \left(\frac{g^{\mu\beta}g_{\mu\beta,\lambda}}{2}\right)\left(\frac{g^{\lambda\sigma}g^*_{\sigma\alpha}}{2}\right) \\
- \quad & \left(\frac{g^{\lambda\beta}g_{\beta\mu,\alpha}}{2}\right)\left(\frac{g^{\mu\sigma}g^*_{\sigma\lambda}}{2}\right) - \frac{g^{44}g_{44,\beta}}{2}\left(\frac{g^{\beta\lambda}g^*_{\lambda\alpha}}{2} - \frac{\delta^{\beta}_{\alpha}g^{\mu\nu}g^*_{\mu\nu}}{2}\right) \quad . \qquad (2.56)
\end{aligned}
$$

Noting that $\partial/\partial x^a = \delta^{\beta}_{\alpha}(\partial/\partial x^{\beta})$ and that $-g^{44}g_{44,\beta}/2 = \sqrt{g_{44}}(\partial/\partial x^{\beta})(1/\sqrt{g_{44}})$ allows us to obtain finally

$$
\begin{aligned}
\frac{R_{4\alpha}}{\sqrt{g_{44}}} \quad = \quad & \frac{\partial}{\partial x^{\beta}}\left[\frac{1}{2\sqrt{g_{44}}}\left(g^{\beta\lambda}g^*_{\lambda\alpha} - \delta^{\beta}_{\alpha}g^{\mu\nu}g^*_{\mu\nu}\right)\right] \\
+ \quad & \left(\frac{g^{\mu\beta}g_{\mu\beta,\lambda}}{2}\right)\left(\frac{g^{\lambda\sigma}g^*_{\sigma\alpha}}{2\sqrt{g_{44}}}\right) - \left(\frac{g^{\lambda\beta}g_{\beta\mu,\alpha}}{2}\right)\left(\frac{g^{\mu\sigma}g^*_{\sigma\lambda}}{2\sqrt{g_{44}}}\right) \quad . \qquad (2.57)
\end{aligned}
$$

This form suggests we should introduce the 4-tensor

$$
P^{\beta}_{\alpha} \equiv \frac{1}{2\sqrt{g_{44}}}\left(g^{\beta\lambda}g^*_{\lambda\alpha} - \delta^{\beta}_{\alpha}g^{\mu\nu}g^*_{\mu\nu}\right) \quad . \qquad (2.58)
$$

The divergence of this is

$$
P^{\beta}_{\alpha;\beta} = (P^{\beta}_{\alpha})_{,\beta} + \Gamma^{\beta}_{\beta\mu}P^{\mu}_{\alpha} - \Gamma^{\mu}_{\alpha\beta}P^{\beta}_{\mu} \quad ,
$$

which when written out in full may be shown to be the same as the r.h.s. of (2.57). The latter therefore reads

$$
\frac{R_{4\alpha}}{\sqrt{g_{44}}} = P^{\beta}_{\alpha;\beta} \quad . \qquad (2.59)
$$

The field equations (2.49) as $R_{4\alpha} = 0$ can then be summed up by the relations

$$
\begin{aligned}
P^{\beta}_{\alpha;\beta} \quad &= \quad 0 \quad , \\
P^{\beta}_{\alpha} \quad &\equiv \quad \frac{1}{2\sqrt{g_{44}}}\left(g^{\beta\sigma}g^*_{\sigma\alpha} - \delta^{\beta}_{\alpha}g^{\mu\nu}g^*_{\mu\nu}\right) \quad . \qquad (2.60)
\end{aligned}
$$

These have the appearance of conservation laws for P_α^β. The fully covariant form and associated scalar for the latter are:

$$\begin{aligned} P_{\alpha\beta} &= \frac{1}{2\sqrt{g_{44}}}\left(g^*_{\alpha\beta} - g_{\alpha\beta}g^{\mu\nu}g^*_{\mu\nu}\right) \\ P &= \frac{-3g^{\lambda\sigma}g^*_{\lambda\sigma}}{2\sqrt{g_{44}}} \quad . \end{aligned} \tag{2.61}$$

We will examine these quantities elsewhere, but here we comment that while our starting gauge (2.34) removed the explicit electromagnetic potentials, the field equations $R_{4\alpha} = 0$ or (2.60) are of electromagnetic type.

It is apparent from the working in this section that the starting conditions (2.34) provide a convenient way to split the 5D field equations $R_{AB} = 0$ into 3 sets: The 5D equations $\hat{R}_{\alpha\beta} = 0$ give a set of equations in the 4D Ricci tensor $R_{\alpha\beta}$ (2.50); the 5D equation $R_{44} = 0$ gives a wave-like equation in the scalar potential (2.51); and the 5D equations $R_{4\alpha} = 0$ can be expressed as a set of 4D conservation laws (2.60). Along the way we also obtain some other useful relations, notably an expression for the 4D Ricci scalar in terms of the dependency of the 4D metric on the extra coordinate (2.52). However, the physically most relevant expression is an effective or induced 4D energy-momentum tensor (2.53). Another way to express these results is to say that the 15 field equations $R_{AB} = 0$ of (2.1) or $G_{AB} = 0$ of (2.2) can always be split into 3 sets which make physical sense provided the metric is allowed to depend on the extra coordinate x^4. These sets consist of 4 conservation equations of electromagnetic type, 1 equation for the scalar field of wave type, and 10 equations for fields and matter of gravitational type. In fact, the last *are* Einstein's equations (2.3) of general relativity, with matter induced from the extra dimension.

2.6 Conclusion

The idea of embedding $G_{\alpha\beta} = 8\pi T_{\alpha\beta}$ (4D) in $R_{AB} = 0$ (5D) is motivated by the wish to *explain* classical properties of matter rather than merely accepting them as given. In application to the cosmological case it works straightforwardly, and gives back 5D geometric quantities which are identical to the 4D density and pressure (Section 2.3). This is important: what we derive from the 5D equations is not something esoteric but ordinary matter. In application to the soliton or 1-body case, the idea leads to a class of radiation-like solutions which contains as a very special case the Schwarzschild solution (Section 2.4). In general application to neutral matter, the properties of the latter turn out to be intimately connected to x^4-dependency of the metric (Section 2.5). Induced-matter theory actually admits a wide variety of equations of state (Ponce de Leon and Wesson 1993). But in the matter gauge at least, independence from x^4 implies radiation-like matter, while dependence on x^4 implies other kinds of matter.

The theoretical basis we have demonstrated in this chapter leads naturally to the question of observations, particularly with regard to the solitons. As mentioned above, there is a class of these in 5D rather than the unique Schwarzschild solution of 4D, because Birkhoff's theorem in its conventional form does not apply. Indeed, there are known exact solutions which represent time-dependent solitons (Liu, Wesson and Ponce de Leon 1993; Wesson, Liu and Lim 1993). And there is known an exact solution which is x^4-dependentand Schwarzschild-like (Mashhoon, Liu and Wesson, 1994). We will return to the latter, where we will find that it implies the same dynamics as in general relativity and so poses no problem. However, there remains the question of the observational status of the standard solitons. This has been investigated by a number of people, most of whom were not working in the induced-matter picture (see Overduin and Wesson 1997). Here, we can regard the soliton as a concentration of matter at the centre of ordinary space, and ask about the motions of test particles at large distances. Specifically, we ask what constraints we can put on the soliton 1-body metric from the classical tests of relativity.

Bibliography

[1] Abolghasem, G., Coley, A.A., McManus, D.J. 1996, J. Math. Phys. 37, 361.

[2] Billyard, A., Wesson, P.S., Kalligas, D. 1995, Int. I. Mod. Phys. D 4, 639.

[3] Billyard, A., Wesson, P.S. 1996, Gen. Rel. Grav. 28, 129.

[4] Brout, R., Englert, F., Gunzig, E. 1978, Ann. Phys. (New York) 115, 78.

[5] Campbell, J.E. 1926, A Course of Differential Geometry (Clarendon, Oxford).

[6] Chattejee, S. 1990, Astron. Astrophys. 230, 1.

[7] Chattejee, S., Sil, A. 1993, Gen. Rel. Grav. 25, 307.

[8] Chattejee, S., Panigrahi, D., Banerjee, A. 1994, Class. Quant. Grav. 11, 371.

[9] Davidson, A., Owen, D. 1985, Phys. Lett. B 155, 247.

[10] Eisenhart, L.P. 1949, Riemannian Geometry (Princeton U. Press., Princeton).

[11] Gross, D.J., Perry, M.J. 1983, Nucl. Phys. B 226, 29.

[12] Guth, A.H. 1981, Phys. Rev. D 23, 347.

[13] Kramer, D., Stephani, H., MacCallum, M., Herlt, E. 1980, Exact Solutions of Einstein's Field Equations (Cambridge U. Press, Cambridge).

[14] Lidsey, J.E., Romero, C., Tavakol, R., Rippl, S. 1997, Class. Quant. Grav. 14, 865.

[15] Liu, H., Wesson, P.S., Ponce de Leon, J. 1993, J. Math. Phys. 34, 4070.

[16] Liu, H., Wesson, P.S. 1994, Int. J. Mod. Phys. D 3, 627.

[17] Liu, H., Mashhoon, B. 1995, Ann. Phys. (Leipzig) 4, 565.

[18] Mashhoon, B., Liu, H., Wesson, P.S. 1994, Phys. Lett. B 331, 305.

[19] Overduin, J., Wesson, P.S. 1997, Phys. Rep. 283, 303.

[20] Ponce de Leon, J. 1988, Gen. Rel. Grav. 20, 539.

[21] Ponce de Leon, J., Wesson, P.S. 1993, J. Math. Phys. 34, 4080.

[22] Rippl, S., Romero, C., Tavakol, R. 1995, Class. Quant. Grav. l2, 2411.

[23] Romero, C., Tavakol. R., Zalaletdinov, R. 1996, Gen. Rel. Grav. 28, 365.

[24] Schouten, J.A., Struik, D.J. 1921, Am. J. Math. 43, 213.

[25] Sorkin, R.D. 1983, Phys. Rev. Lett. 5l, 87.

[26] Tangherlini, F.R. 1963, Nuovo Cim. 21 (Ser. 10), 636.

[27] Wesson, P.S. 1992a, Astrophys. J. 394, 19.

[28] Wesson, P.S. 1992b, Phys. Lett. B 276, 299.

[29] Wesson, P.S., Ponce de Leon, J. 1992, J. Math. Phys. 33, 3883.

[30] Wesson, P.S., Liu, H., Lim, P. 1993, Phys. Lett. B 298, 69.

[31] Wesson, P.S. 1994, Astrophys. J. 436, 547.

[32] Wesson, P.S., Ponce de Leon, J. 1994, Class. Quant. Grav. ll, 1341.